MEASURING STANDARDISED MORTALITY RATIOS OF HOSPITALS

Challenges and recommendations

Measuring standardised mortality ratios of hospitals
Challenges and recommendations

ISBN 978-90-393-66776
Design Jolien van der Wilden
Print Ten Brink

MEASURING STANDARDISED MORTALITY RATIOS OF HOSPITALS

Challenges and recommendations

Meten van gestandaardiseerde sterftecijfers van ziekenhuizen

Uitdagingen en aanbevelingen
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op dinsdag 29 november 2016 des middags te 2.30 uur

door

Maurice Emir Pouw

geboren op 5 september 1976 te Delft

Promotoren:
Prof. Dr. C.J. Kalkman
Prof. Dr. K.G.M. Moons

Copromotor:
Dr. L.M. Peelen

To my parents,
Hway In Pouw and Bie Kiok Pouw-Tan

CONTENTS

CHAPTER 1

General introduction

GENERAL INTRODUCTION

Hospital mortality is considered one of the key measures to assess and compare quality of care of hospitals.[1,2] It is however difficult to interpret mortality figures as they are influenced by numerous factors, which are inseparable and often unrelated to *delivered quality* of hospital care. In this, one of the key issues is the fact that patient populations differ between hospitals, so called casemix differences. The hospital standardised mortality ratio (HSMR), developed by Jarman et al., aims to adjust for those casemix differences between hospitals in an attempt to benchmark and judge in-hospital mortality numbers.[3] The HSMR compares the number of in-hospital deaths of a hospital with an expected number of deaths for that hospital as computed with a casemix correction model. As a consequence, a commonly made assumption is that greater than expected mortality reflects 'avoidable' deaths and thus indicates lesser quality of care.[1]

In numerous countries, including The Netherlands, HSMRs are made public with the aim of transparency and 'learning from benchmarking'. However, since its introduction several concerns have been raised regarding the calculation of the HSMR and its ability to reflect quality of hospital care.[4-9] Van Gestel et al. described and summarised the most important shortcomings and grouped them into 'disease severity', 'referral bias', 'place of death and end-of-life care' and 'casemix and coding' issues.[10] Making invalid measures public can have devastating consequences for reputations of hospitals, relationships between health care providers, and hospitals' financial status, especially if sanctions or rewards are linked to these measures. Various authors claim that the shortcomings of HSMR calculation are such that the use of HSMR as a tool to assess quality of care is incorrect and should not be published.[4,7,10]

As hospital mortality statistics are considered important, attractive and relatively inexpensive to measure, public and policy makers will not easily accept ignoring HSMRs. Because HSMR as a performance indicator for quality of care in hospitals is likely here to stay, it is of utmost importance to gain a clear understanding of both its strengths and shortcomings, and to formulate adjustments to improve the validity of HSMRs.

Objectives of this thesis

This thesis has the following objectives:
- To study the effects of 'referral bias' and 'casemix and coding issues' on the current Dutch HSMR calculation.
- To identify potential adjustments in the estimation of the HSMR to improve its validity as a performance indicator.

Outline of this thesis

The thesis starts with investigating the theoretical method underlying the calculation of the HSMR, the so-called indirect standardisation method. In chapter 2, the indirect standardisation method is compared with the direct standardisation method. Also, pitfalls of HSMR resulting from the indirect standardisation method are discussed, and recommendations are given to reduce the shortcomings of this method.

Subsequently, the thesis investigates potential modifications of the currently used model for HSMR calculation. To adjust for casemix differences between hospitals, parameters of comorbidities are included in the model underlying the HSMR calculation. In chapter 3, the commonly used Charlson comorbidity measure is compared with the Elixhauser comorbidity measure. Discriminative performance of the casemix correction models based on these two comorbidity measures is compared and their effects on the HSMRs of individual hospitals are explored.

The Dutch HSMR is currently based on in-hospital mortality. However, discharge patterns, average length of hospital stay, and transfers all affect in-hospital mortality. In chapter 4, effects of the inclusion of post-discharge mortality on HSMRs are compared with those of in-hospital mortality.

In the final part of the thesis we zoom in onto the mortality ratios of specific patient populations, rather than that of an entire hospital population. In chapter 5, the focus is on SMRs of specific diagnosis groups requiring specialised care offered by specialised hospitals. The SMRs of specialised and non-specialised hospitals are compared and the influence of referral patterns on SMRs is investigated.

Current HSMR calculation is based on administrative databases and said to lack important clinical predictors. In chapter 6, the casemix adjustment model for cardiac surgery patients, based on an administrative database, is compared with the validated clinical EuroSCORE prediction model, based on a clinical database. Also influences of the two models on eventual SMRs are compared.

Finally, in chapter 7, the results and implications of this thesis are summarised and discussed together with insights and recommendations to improve the validity and utility of HSMRs.

REFERENCE LIST

1 www.igz.nl/zoeken/download.aspx?download=De+veiligheid+telt.pdf
2 https://www.nza.nl/1048076/1048133/NR_CU_244__Regeling_verplichte_publicatie_sterftecijfers_instellingen_voor_medisch_specialistische.pdf
3 Jarman B, Galt S, Alves B, Hider A, Dolan S, Cook A, et al. Explaining differences in English hospital death rates using routinely collected data. *BMJ* 1999; 318:1515-20.
4 Lilford R, Pronovost P. Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *BMJ* 2010; 340:c2016.
5 Penfold RB, Dean S, Flemons W, et al. Do hospital standardized mortality ratios measure patient safety? HSMRs in the Winnipeg Regional Health Authority. *Healthc Pap* 2008; 8:8-24; discussion 69-75.
6 Mohammed MA, Deeks JJ, Girling A, et al. Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals. *BMJ* 2009; 338:b780
7 Pitches DW, Mohammed MA, Lilford RJ. What is the empirical evidence that hospitals with higher-risk adjusted mortality rates provide poorer quality care? A systematic review of the literature. *BMC Health Serv Res* 2007; 7:91
8 Van den Bosch WF, Roozendaal KJ, Silberbusch J. Fluctuating mortality rates. HSMR no reliable measure for quality of care. *Medisch Contact* 2009; 64:31-2
9 Goodacre S, Campbell M, Carter A. What do hospital mortality rates tell us about quality of care? *Emerg Med J* 2015; 32:244-7
10 Van Gestel YRBM, Lemmens VEPP, Lingsma HF, et al. The Hospital Standardized Mortality Ratio Fallacy. A Narrative Review. *Med Care* 2012; 50:662-7

CHAPTER 2

Hospital standardised mortality ratio: consequences of adjusting hospital mortality with indirect standardisation

Pouw ME, Peelen LM, Lingsma HF, Pieter D, Steyerberg E, Kalkman CJ, Moons KGM

PLoS One 2013; 8(4): e59160

ABSTRACT

Background

The hospital standardised mortality ratio (HSMR) is developed to evaluate and improve hospital quality. Different methods can be used to standardise the hospital mortality ratio. Our aim was to assess the validity and applicability of directly and indirectly standardised hospital mortality ratios.

Methods

Retrospective scenario analysis using routinely collected hospital data to compare deaths predicted by the indirectly standardised casemix adjustment method with observed deaths. Discharges from Dutch hospitals in the period 2003-2009 were used to estimate the underlying prediction models. We analysed variation in indirectly standardised hospital mortality ratios (HSMRs) when changing the casemix distributions using different scenarios. Sixty-one Dutch hospitals were included in our scenario analysis.

Results

A numerical example showed that when interaction between hospital and casemix is present and casemix differs between hospitals, indirectly

standardised HSMRs vary between hospitals providing the same quality of care. In empirical data analysis, the differences between directly and indirectly standardised HSMRs for individual hospitals were limited.

Conclusion

Direct standardisation is not affected by the presence of interaction between hospital and casemix and is therefore theoretically preferable over indirect standardisation. Since direct standardisation is practically impossible when multiple predictors are included in the casemix adjustment model, indirect standardisation is the only available method to compute the HSMR. Before interpreting such indirectly standardised HSMRs the casemix distributions of individual hospitals and the presence of interactions between hospital and casemix should be assessed.

INTRODUCTION

In the last decades increasing attention is directed towards the quality of care of hospitals. Various performance indicators have been developed to express quality of care, among which the hospital standardised mortality ratio (HSMR). The HSMR is a risk adjusted hospital mortality rate that corrects crude hospital mortality rates by taking into account the casemix of the hospital.[1] Developed and implemented in 1999, the HSMR is now used as a key hospital quality indicator in various countries including the United Kingdom, the United States, Canada, and the Netherlands.[2-7]

The HSMR is used by hospitals, health authorities, and media as a tool to assess the delivered quality of care, to analyse the trend of the quality of care of a hospital over time, and to compare and rank hospitals. Since its introduction, the HSMR has been debated for various reasons: the credibility of the link between quality of care and risk adjusted mortality,[8-10] the variables that are used for casemix adjustment,[11] and issues regarding coding of these variables.[12]

Another important, but often neglected issue, is the fact that the HSMR is computed via the so-called *indirect standardisation* method. It has been long known that if mortality rates are adjusted via the indirect standardisation method, these rates cannot always be compared.[13-16] However, it seems almost inevitable that HSMRs of hospitals will be compared and 'quality performance league tables' will be constructed.

The present paper illustrates the potential pitfalls of HSMR when used to compare hospitals. We will first provide a description of the indirect and direct standardisation method to demonstrate why caution must be taken when hospitals are compared and ranked based on indirectly standardised figures like the HSMR. Subsequently, we illustrate the consequences of indirect standardisation in practice using HSMR figures from the Netherlands.

METHODS

Ethics statement

To study the impact of this phenomenon caused by indirect standardisation on real clinical data, we have conducted a series of analyses on the Dutch HSMR figures, permitted by the Dutch Hospitals Association and the Dutch University Medical Centers Association. The data were obtained from the Dutch National Medical Registration database, which contains routinely collected hospital episode statistics of Dutch patients and is held by Dutch Hospital Data. All data were analyzed anonymously (http://www.dutch-hospitaldata.nl/Bestanden/Documenten/Protocol_gegevensgebruik_DHD_databanken.pdf).

Standardisation methods

Differences in crude mortality rates between hospitals are not only caused by differences in hospital performance but also by differences in the casemix of patients that are admitted. A hospital that admits on average older patients and performs a larger proportion of 'high risk' procedures is likely to have a higher in-hospital mortality rate than a hospital with on average younger patients and a smaller proportion of 'high risk' procedures. Standardisation methods use information at patient level such as reason of admission, age, sex, deprivation category and comorbidity to adjust for these differences in casemix.

A standardised mortality ratio is calculated as the observed number of deaths divided by the expected number of deaths. For the HSMR, this is the observed and expected mortality for a given hospital in a given year, expressed as a percentage. If the observed number of deaths is 120 and the expected number of deaths is 100, the HSMR for that hospital would be 120. A HSMR greater than 100 reflects more deaths than expected and a HSMR less than 100 reflects fewer deaths than expected.

There are two main methods of standardisation: direct and indirect. The main difference between these two methods is *what* is being standardised, whether it being the casemix (direct standardisation) or the mortality rate (indirect standardisation).

Direct standardisation The direct standardisation method standardises the casemix of patients admitted in a hospital to a *reference* casemix. A directly standardised mortality rate of a hospital is therefore based on the same casemix as the directly standardised mortality rates of other hospitals, i.e. on the *reference population (reference casemix)*. In this way the effect of differences in casemix populations between hospitals is eliminated.

Directly standardised mortality rates are computed as follows. First, the probability of in hospital death is calculated for each subcategory of patients as number of deaths divided by the number of admissions in that subcategory. Thus for example the probability of in hospital death for men, treated for the diagnosis of pneumonia, in the age category 60-64 year, may be 2% in one hospital, whereas patients with the same combination of predictors may have a mortality probability of 3% in another hospital.

Secondly, the mortality probabilities of each hospital are applied to the same reference hospital population to obtain the expected number of deaths *in the reference hospital population*. If the reference hospital population has 100 patients in the subcategory of our example, the expected number of deaths according to the mortality rates of our example hospital would be 2 (100 x 2%). Summation of the expected number of in-hospital deaths of all subcategories gives the total expected number of in-hospital deaths in the reference hospital population. The ratio between the expected number of in-hospital deaths and the actual number of in-hospital deaths *in the reference population* gives the directly standardised mortality ratio for the hospital of interest. Note that a hospital must have patients in a subcategory to calculate the corresponding mortality rate for that subcategory, and that the number of patients in the subcategory must be large enough to obtain reliable mortality rates.

Indirect standardisation The indirect standardisation method standardises the mortality rate of the casemix to a *reference mortality rate* (expected mortality rate). An indirectly standardised mortality rate of a hospital is based on the expected mortality rate for that hospital given its casemix of patients.

The indirect standardisation method calculates the expected number of deaths for a hospital in two steps. First, *expected* probabilities of in-hospital death are computed using a logistic regression model with in-hospital mortality (yes/no) as outcome and various patient characteristics as predictors. For this modelling one commonly uses the data of many (preferably all) hospitals in a particular country. These expected probabilities of in-hospital death are computed for each subcategory of patients and can be interpreted

as the probability of in-hospital death for patients belonging to the corresponding subcategory in a *standard* hospital of that country. For example the prediction model might calculate that the expected in-hospital probability of death for men, treated for the diagnosis of pneumonia, in the age category 60-64 year is 3%.

Secondly, these expected probabilities are applied to the admission numbers of a specific hospital to compute the expected number of deaths *in that hospital*. If the hospital under study has admitted 200 patients in the subcategory of our example, the expected number of deaths in this subcategory would be 3% of 200 or 6 deaths. The summation of the expected number of deaths in all subcategories gives the total expected number of deaths for that hospital. The observed number of deaths in a hospital is calculated by simply counting the number of people who died *in the specific hospital* within the given period. The ratio between the observed number of deaths and the expected number of deaths gives the indirectly standardised mortality ratio.

The advantage of directly standardised mortality rates is that these rates are comparable with each other because the effect of differences in casemix is eliminated, as they are all based on the same reference hospital population. However, a subcategory of patients of a hospital under study may be very small, resulting in an unreliable mortality rate (e.g. a mortality rate of 0% in a subcategory containing 5 patients). Moreover, if a hospital does not have any patients in a subcategory, the direct standardisation method cannot be used at all. Therefore, in most cases, direct standardisation is not applicable and the indirect standardisation method is used. This is also the case for the HSMR as it is generally calculated. The drawback of indirectly standardised mortality rates is that these rates are not always comparable with each other as will be explained in the example below.

Numerical example

We assume there are two hospitals, hospital A and hospital B, of equal size (both admitting 5000 patients per year) and both delivering precisely the same quality of care. If the HSMR is a fair and valid measure of quality of care, this measure should then also be equal for both hospitals. For reasons of simplicity, we distinguish only two kinds of patients (i.e. using one patient characteristic instead of the 9 normally used to calculate the expected mortality): urgently versus non-urgently admitted patients. Suppose that hospital A has admitted 20% of the 5000 patients urgently and hospital B 80%.

We now assume that urgently admitted patients have an *expected* mortality rate of 6% and non-urgently admitted patients of 2% (see also table 2.1). Furthermore, assume that in both hospitals the *observed* mortality rates for these groups are 3% and 4% respectively. Although a patient admitted to hospital A has the same chance to die as in hospital B (3% if admitted urgently and 4% if admitted non-urgently) the HSMR is 136 for hospital A and 61 for hospital B (table 2.1).

Table 2.1 Numerical example of direct and indirect standardisation

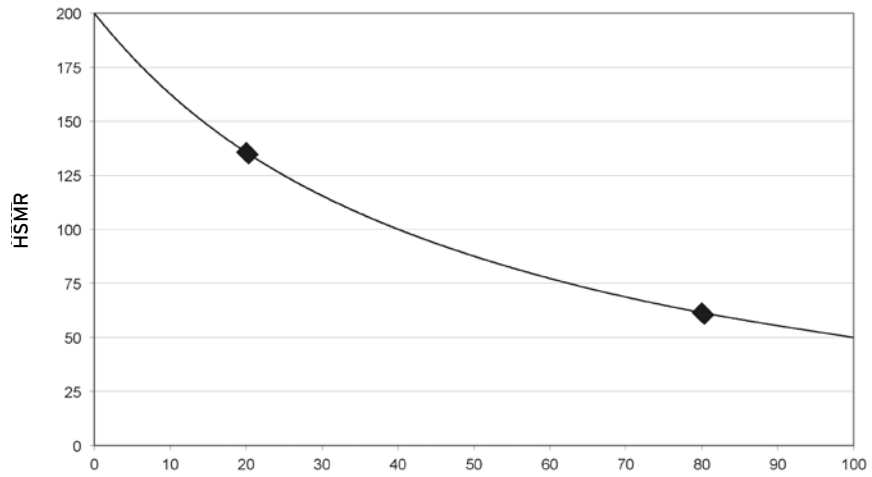
	Hospital A		Hospital B	
	Urgent	Non-urgent	Urgent	Non-urgent
Expected mortality rate	6%	2%	6%	2%
Observed mortality rate	3%	4%	3%	4%
Casemix	1000	4000	4000	1000
Indirect standardisation	**136** 1000 x 3% + 4000 x 4% 1000 x 6% + 4000 x 2%		**61** 4000 x 3% + 1000 x 4% 4000 x 6% + 1000 x 2%	
Direct standardisation	**100** 1000 x 3% + 4000 x 4% 1000 x 3% + 4000 x 4%		**100** 1000 x 3% + 4000 x 4% 1000 x 3% + 4000 x 4%	

Although both hospitals have the same observed mortality hospital A performs worse than hospital B when the mortality rate is adjusted via the indirect standardisation method.

The difference in indirectly standardised HSMR is the result of the difference in casemix and of interaction between casemix and hospital and is explained as follows. In both hospitals, urgently admitted patients have a lower probability to die than non-urgently admitted patients (observed rates 3% vs. 4%). In the total population the effect of urgency is the other way around (expected rates 6% vs. 2%). This means that there is statistical interaction between hospital and urgency. Although the two hospitals in the example perform similar, hospital B benefits from this situation as the majority of its population consists of urgently admitted patients (80% of the total), resulting in a lower HSMR than for Hospital A. Thus, when comparing and ranking the

performance based on the single HSMR statistic only, hospital B is considered to be better than hospital A, despite the fact that the chance to die for a random patient is equal in both hospitals. Figure 2.1 displays the relation between the casemix distribution (urgent – non-urgent ratio) and the HSMR, keeping the expected and observed mortality rates constant.

Figure 2.1 Change in HSMR when the ratio of urgently vs. non-urgently admitted patients changes



The observed mortality rates are 3% and 4% for respectively urgently and non-urgently admitted patients.The expected rates are 6% and 2%, implying the presence of statistical interaction between hospital and urgency, which is ignored in the adjustment model. Markers indicate the proportions of urgently admitted patients used in the theoretical example (20% and 80% respectively).

The direct standardisation method is not affected by this present interaction. Suppose hospital A is the reference hospital. Then the casemix of hospital B will be standardised to the casemix of hospital A. Because the observed mortality rates for urgently and non-urgently admitted patients do not differ between the hospitals, the directly standardised mortality rates for hospital B and for hospital A are both 100 (see table 2.1).

Besides the use of the HSMR for comparisons across hospitals, it is also advocated to compare the HSMRs of a single hospital over time as an indicator of change in quality of care. However, the same phenomenon as described above can be found. When the casemix distribution changes over time and interaction between hospital and casemix is present, the HSMR can still

change even if the quality of care (expressed as observed mortality rates) and the predicted risk for each patient remains constant. A worked-out example can be found in Appendix S1.

Application to Dutch HSMR figures

To study the impact of this phenomenon caused by indirect standardisation on real clinical data, we have conducted a series of analyses on the Dutch HSMR figures, permitted by the Dutch Hospitals Association and the Dutch University Medical Centers Association. For the present analyses, patient consent was not necessary as the data was stored and thus used completely anonimized. For the same reason, approval of a medical ethics committee was not needed.

The Dutch HSMR have been calculated in a similar manner to that used in several other countries and was performed by the authors (notably DP) in close collaboration with Dr Foster Intelligence (London, UK). For a detailed description of the Dutch HSMR models and used method, we refer to a previous publication 2. In short: 50 diagnostic groups were selected which accounted for 80% of in-hospital mortality. For each diagnostic group a prediction model (logistic regression model) was fitted using various predictors, including age, gender, urgency of admission, month of admission, Charlson Comorbidity Index, diagnosis, and social deprivation, to generate an expected mortality risk for each admitted patient. In total, 4,031,829 admissions in the period 2003-2009 were included to fit these 50 prediction models. The HSMR is the sum of the observed mortalities in all 50 diagnostic groups divided by the sum of all expected mortalities. The coefficients of the predictors of the final models can be shown on request.

For the present analysis, we tested whether there was an interaction between hospital and urgency of admission, as we hypothesized that the effect of the variable 'urgency of admission' on the outcome (death) might differ across hospitals. For example, high-level trauma centres are probably more adequate in treating acutely admitted patients. For each of the 50 diagnostic groups, we fitted a logistic regression model with the variables 'hospital', 'urgency of admission' and their interaction term 'hospital*urgency of admission' as predictors and tested whether the interaction term was significant ($P<0.05$). We repeated this analysis to test for interaction between hospital and comorbidity.

We analysed the HSMR of 61 Dutch hospitals from the period 2006-2009. For each hospital, we first calculated the HSMR according to the regular indirect standardisation method. Then we analysed eight scenarios.

In the first part (scenario 1-4) we stratified patients according to their admission status into urgent and non-urgent admissions to mimic the numerical example as described above. For both groups, the size of the group, the observed mortality rate, and the expected mortality rate were calculated. Subsequently, we kept the observed and expected mortality rates for these two groups constant, and replaced the original distribution of urgent and non-urgent admissions for each hospital by the 'average casemix distribution of all 61 Dutch hospitals' (scenario 1). The obtained HSMRs reveal what the HSMR of the hospital would have been if that hospital had an average Dutch hospital distribution of urgent and non-urgent admissions. To investigate more extreme variations, we extended these scenarios by replacing the original distribution of a single hospital by the casemix distribution of a single other hospital (scenario 2).

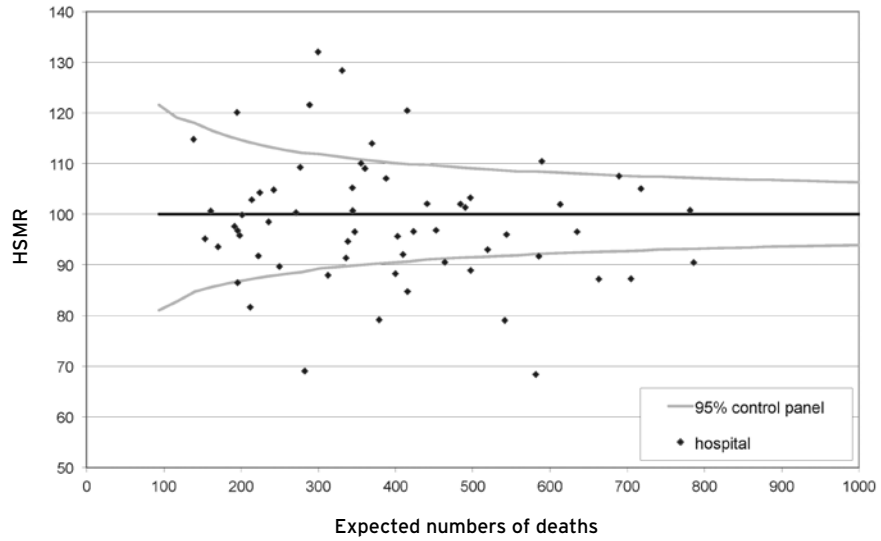
In a third scenario we looked at the effect of differences in casemix distributions of a single hospital on the HSMR over time. We used the observed and expected mortality rates of the urgent and non-urgent admissions of a hospital in the year 2009 as a basis. For each hospital, we recalculated the HSMR for 2009 using the hospital's average casemix distribution of urgent versus non-urgent admissions over the years 2006 – 2009 (scenario 3). Finally, we recalculated the HSMR of each hospital with the distribution of urgent versus non-urgent admissions of the years 2006, 2007 and 2008 separately. Here, differences in HSMRs are then solely to be attributed to differences in distribution between urgent and non-urgent admissions over time (scenario 4).

We repeated these scenario studies using another casemix variable 'Charlson Comorbidity index' (CCI) instead of 'urgency of admission' (scenario 5-8). The CCI is used as a score for comorbidity and is based on 17 comorbidities such as cancer, congestive heart failure, cerebral vascular disease, peripheral vascular disease, dementia, diabetes, and renal disease.[17,18] Each comorbidity is assigned a weighted score. Depending on the patient's sort and number of comorbidities the CCI stratifies the patient into a class ranging from 0 (no comorbidity) to 6 (severe comorbidity).

RESULTS

Figure 2.2 shows a funnel plot of the HSMRs of the 61 hospitals. The funnel plot divides hospitals in three categories using 95% control limits. The 95% control limits demarcate the 95% confidence interval of the HSMR given the expected mortality. Hospitals above the 95% control limits have a HSMR significantly higher than 100, hospitals below the 95% control limits have a HSMR significantly lower than 100, and for hospitals between the 95% control limits a deviation from the reference value of 100 is considered to be a result of natural random variation. As can be seen in figure 2.2, in 2009 the HSMRs of Dutch hospitals differed considerably. Fifteen hospitals appeared to perform significantly better, and seven hospitals significantly worse than expected. According to the risk adjustment model used, the risk of dying in the hospital with the lowest HSMR is 1.93 times lower than the hospital with the highest HSMR (132/68).

Figure 2.2 Funnel plot showing the HSMRs of Dutch hospitals in 2009



In 2009 the 61 hospitals had 492,099 admissions of which 301,916 admitted urgently. Figure 2.3 illustrates the percentage of urgently admitted patients per hospital in 2009, which ranged from 38% to 76% (median 65% (IQR: 60% – 68%). Respectively 53%, 20%, 15%, 5%, 1%, 5%, and 1% of the studied admissions were classified in the CCI group 0, 1, 2, 3, 4, 5, and 6. Figure 2.4 illustrates the distribution of the CCI of admissions per hospital in 2009.

Figure 2.3 Proportion of urgently admitted patients per hospital

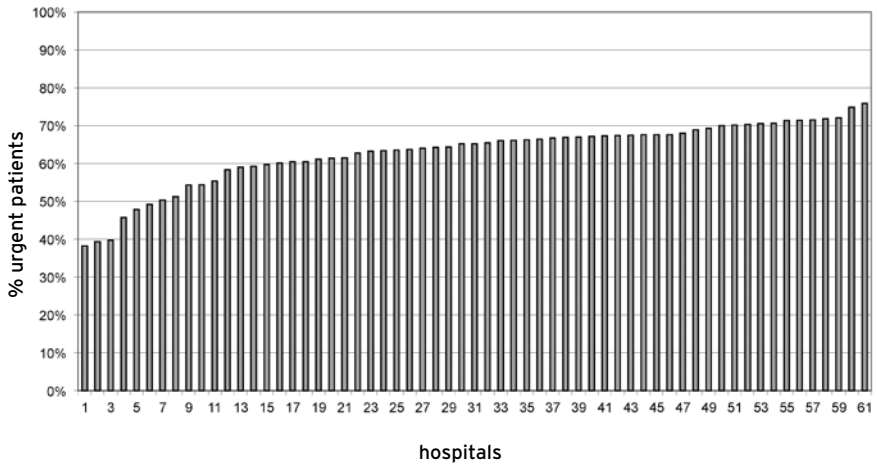
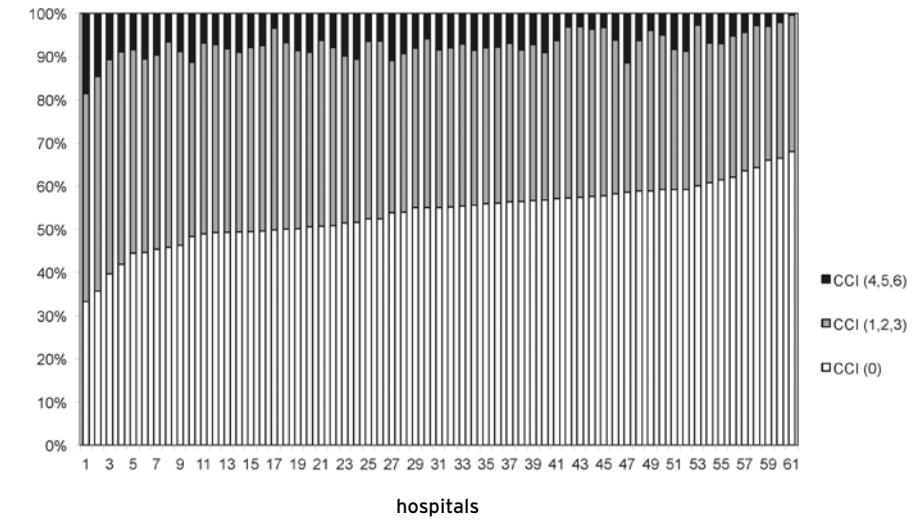


Figure 2.4 Distribution of the Charlson Comorbidity Index of patients per hospital



The Charlson Comorbidity Index groups 1, 2 and 3 are aggregated as well as the groups 4, 5 and 6.

We tested for interaction between hospitals and two casemix variables: 'urgency of admission' and CCI. We found evidence of interaction between hospitals and 'urgency of admission' in 19 of the 50 prediction models (statistically significant interaction term, $P < 0.05$). In 7 of the 50 prediction models we found evidence of interaction between hospitals and CCI.

For each hospital we recalculated the HSMR using the mean of the distributions of the casemix variable 'urgency of admission' of the 61 hospitals (scenario 1). The relative change between these obtained simulated HSMR and the original HSMR ranged from -2.6% for one hospital (HSMR: 110, simulated HSMR: 107) to 5.5% for another hospital (HSMR: 79, simulated HSMR: 83). No hospital changed significantly from category (i.e. the same fifteen and seven hospitals respectively over- and underperformed according to the new computed HSMR). Appendix S2 shows the HSMR and the HSMR of scenario 1 for all hospitals. The top 5 and bottom 5 hospitals remained unchanged when the HSMR of scenario 1 was used to rank the hospitals (Table 2.2).

Table 2.2 Top and bottom 5 hospitals based on their HSMRs

Original		Scenario 1			Scenario 5		
Ranking	HSMR	HSMR-1	New ranking	Rank difference	HSMR-5	New ranking	Rank difference
1	68	69	2	-1	69	1	0
2	69	68	1	1	73	2	0
3	79	83	5	-2	77	4	-1
4	79	79	3	1	76	3	1
5	82	82	4	1	82	5	0
57	120	121	58	-1	114	53	4
58	120	120	57	1	117	54	4
59	122	125	59	0	123	56	3
60	128	128	60	0	129	59	1
61	132	132	61	0	155	61	0

The HSMR of scenario 1 is computed based on the mean of the casemix distributions of the 'urgency of admission' variable of the 61 hospitals. The HSMR of scenario 5 is computed based on the mean of the casemix distributions of the 'Charlson Comorbidity index' variable of the 61 hospitals. Hospitals ranked in top 5 were all significantly lower than 100, whereas the bottom 5 hospitals were all significantly higher than 100.

The total absolute difference in HSMR was 44 points (an average of 0.72 points per hospital). Tables 3.3 and 3.4 show an overview of the results of the eight scenarios. In scenario 2, where we replaced the casemix distribution of a hospital with the distribution of one single other hospital, we found that for 10 (16.4%) hospitals another hospital could be found whose casemix distribution significantly changed the category in the funnel plot. For 7 of these 10 hospitals, the HMSR was close to a control limit (less than 2 HSMR points). In scenario 3 we replaced the casemix distribution of a hospital in 2009 with the average casemix distribution of that hospital (2006-2009). No hospital changed significantly from category with the new HSMR. The difference between the HSMR of scenario 3 and the original HSMR ranged from -4.0% for one hospital (the HSMR decreased from 69 to 66) to 0.8% for another hospital (HSMR of 79 increased to 80). The total absolute difference in HSMR was 19 points (an average of 0.32 points per hospital). In scenario 4 we replaced the casemix distribution of a hospital in 2009 with the distribution of that hospital of a single previous year (2006, 2007 and 2008). For one hospital the distribution of year 2008 significantly changed the category in the funnel plot (from better than expected to average). However, the HSMR only changed one point (from 86 to 87). No hospital changed categories when the distribution of 2007 was used and one hospital changed categories when the distribution of 2006 was used (from worse than expected, HSMR of 110, to average, HSMR of 108).

When repeating the scenarios based on the Charlson Comorbidity Index (scenarios 5-8), differences between the original and simulated HSMRs increased. The relative change between the obtained simulated HSMR based on the mean co-morbidity distribution and the original HSMR ranged from -6.7% for one hospital (HSMR: 97, simulated HSMR: 90) to 29.5% for another hospital (HSMR: 95, simulated HSMR: 123). Eight hospitals, including the 2 hospitals mentioned above, changed significantly from category when the HSMR of scenario 5 was compared with its original HSMR. For 7 of these 8 hospitals the difference in HSMR ranged from -4 points to 1 point. The hospital with the largest HSMR difference (from 95 to 123) tumbled in the ranking list from position 23 to position 57. The HSMRs of the hospitals are shown in Appendix S2. The top 5 hospitals remained unchanged when the HSMR of scenario 5 was used to rank the hospitals, but the differences were larger in the bottom 5 hospitals.

Looking at changes over time, we found that the relative change between the HSMR with the average distribution of that hospital of the years 2006-2009 and the original HSMR ranged from -18.6% (HSMR dropped from 97 to 79) to 4.8% (HSMR increased from 97 to 101). Three hospitals significantly changed to another category because of a change in HSMR.

Table 2.3 Scenarios using an average hospital casemix distribution

Scenario	Number of hospitals changing category	Number of hospitals changing ranks			
		No rank change	1-5 ranks	5-10 ranks	>10 ranks
1	0	14	47	0	0
3	0	38	22	1	0
5	8 (3)	10	41	5	5
7	3 (1)	14	42	4	1

In scenario 1 and 5 the mean distribution of the casemix variable under study of the 61 hospitals is used to recalculate the HSMR of the hospitals. In scenario 3 and 7 the HSMR of a hospital is recalculated using the mean distribution of the casemix variable under study over time (2006-2009). In the second column the numbers of hospitals are shown for which the recalculated HSMR crosses a 'control limit'. In brackets: the number of hospitals for which the HSMR lies close to a control limit (within 2 HSMR points). Columns 3 to 7 show an overview of rank changes of hospitals based on the recalculated HSMR.

Table 2.4 Scenarios using a unique hospital casemix distribution

Scenario	Number of hospitals changing category	Significant HSMR change				
		By one hospital	By 2-5 hospitals	By more than 5 hospitals	By 1 other year	By more than 1 year
2	10 (7)	2	3	5	N.A.	N.A.
4	2 (2)	N.A.	N.A.	N.A.	2	0
6	35 (13)	6	4	25	N.A.	N.A.
8	3 (1)	N.A.	N.A.	N.A.	0	3

In scenario 2 and 6 the HSMR is recalculated using the distribution of the casemix variable under study of a single hospital. In these scenarios for each hospital 60 HSMRs are recalculated. In scenario 4 and 8 the HSMR is recalculated using the distribution of the casemix variable under study of another year. In these scenarios for each hospital three HSMRs are recalculated. In the second column the numbers of hospitals are shown for which the recalculated HSMR crosses a 'control limit'. In brackets: the number of hospitals for which the HSMR lies close to a control limit (within 2 HSMR points). Columns 3 to 5 show an overview of the number of hospitals whose casemix distribution changes the HSMR of a hospital significantly. Columns 6 and 7 show an overview of the years where the differences in casemix distribution change the HSMR of a hospital significantly.

DISCUSSION

The HSMR is considered to be an important tool in the assessment of quality of care across hospitals as well as for a single hospital over time. In a growing number of countries, hospital board members, media, and public use the HSMR to monitor and compare the quality of care. Given the fact that the HSMR is based on indirect standardisation, it has been known – although in practice largely ignored – that such comparisons are only allowed if the underlying casemix distributions are identical or if there is no interaction between hospitals and casemix variables.[13-16] In this paper we showed the pitfalls of indirect standardisation of the HSMR by means of a numerical example, and in practice by using Dutch clinical data.

With the numerical example we illustrated that when there is interaction between hospital and casemix, the indirectly standardised HSMR is not only determined by the observed and expected mortality rates, but is also related to the distributions of the underlying casemix variables. Thus, caution must be taken not only when interpreting and comparing HSMRs of different hospitals but also when comparing HSMRs of a given hospital over time. When there is no interaction between hospital and casemix, direct and indirect standardisation will lead to the same HSMR, also when there are differences in casemix distribution.

From our empirical studies we learned that although changing the casemix distribution of a hospital results in a different HSMR, differences between the HSMR calculated with the average Dutch casemix distribution of the variable 'urgency of admission' (scenario 1) and the original HSMR, were small. In terms of ranking, we see little movement, and no hospitals moved across categories. However, replacing the hospitals' casemix by that of a single other hospital (scenario 2) led to significant changes in HSMR and rank. Comparing the HSMR of a hospital with the HSMR computed with the average casemix distribution of previous years of that hospital for the variable 'urgency of admission' (scenario 3), showed again only small differences in final hospital ranking. This is probably because the casemix distribution of this variable did not change substantially over the last 4 years for the Dutch hospitals. From these three scenarios we learned that the larger the difference between the 'original' and the 'new' casemix, the larger the change in HSMR will be due to the indirect standardisation method itself.

The findings are also strongly dependent on which casemix variable is investigated. For the variable 'Charlson Comorbidity index', the distribution differences between hospitals cause more discrepancy between the original HSMR and the simulated HSMR (scenario 5) than the simulation study in which the casemix variable 'Urgency of admission' was used (scenario 1). This is probably due to more variability in the distribution of the Charlson Comorbidity index between hospitals. Nevertheless, the ranking of the top and bottom 10 hospitals still hardly changed. Furthermore, for some hospitals the distribution of the variable 'Charlson Comorbidity index' also varied considerably between years (scenario 8).

Some limitations of our study must be taken into account. We only looked at two of the nine casemix variables used in the casemix correction model. Moreover, in our scenario studies we varied only one variable at a time. Despite the high possibility that the distributions of the other casemix variables also differ between hospitals and over the years, we explicitly chose to focus on these two variables because our goal was to provide insight in the fact that casemix differences between Dutch hospitals can influence the HSMR and because these two variables are subject to debate in terms of coding issues.[12] Therefore, distribution differences between hospitals may distort the comparison of HSMRs more than revealed with this study. However, it might be expected that differences in the distribution of other predictors in the adjustment model would have similar effects.

Another limitation is that we only looked at a period of 4 years (2006-2009). Although our study indicates that differences in casemix distributions of hospitals over time do not influence the HSMR noticeably, it is very well possible that over a longer period of time casemix distributions change, such that long term trend monitoring using the HSMR may be misleading.

Due to their indirect standardisation, HSMRs may not automatically be comparable neither across hospitals nor for a single hospital over time, unless the underlying casemix distributions are proportionally the same or when there is no interaction between hospital and casemix. In our data we found evidence of interaction between hospitals and the casemix variables 'urgency of admission' and 'comorbidity'. However, our empirical example showed that when differences in casemix were limited (scenario 1 and 3), the effect on the HSMRs and thus the ranking of the hospitals is limited. Only replacing a hospital's casemix with a very different casemix (scenario 6) led to significant changes in HSMRs. More importantly, direct standardisation is practically impossible when multiple predictors are included in the adjustment model. The numbers of patients in each subcategory then become too small to obtain reliable mortality rates. Furthermore, it has been previously argued that the indirectly standardised HSMR provides the mortality rate from a societal perspective as it is based on the population the hospital actually serves, not the national reference population, while a HSMR based on direct standardisation is more relevant to informing patient choices.[19]

Although still subject to much discussion, HSMR will likely remain as one of the indicators for hospital quality. HSMRs should be interpreted, however, with the greatest caution, due to issues concerning the link between in hospital mortality and quality of care, coding differences between hospitals, insufficient casemix adjustment, and poor data quality. In addition, in this study we have shown that the indirect standardisation method used to compute the HMSR might also distort the interpretation of HSMRs. Therefore, we urge researchers to first investigate the distributions of the underlying casemix variables and assess the presence of interactions between hospital and casemix. A possible solution might be to analyse hospitals within clusters with comparable casemix distributions such as small regional hospitals, large teaching hospitals and academic hospitals. Comparing HSMRs of hospitals belonging to the same cluster reduces the chance that differences in casemix distributions and interaction are the cause of HSMR differences across hospitals. Also for trend monitoring of HSMRs within a single hospital, the casemix distributions of those years, and potential interaction between year and casemix must be analysed before interpreting possible changes in HSMR.

REFERENCE LIST

1 Jarman B, Galt S, Alves B, Hider A, Dolan S, et al. (1999) Explaining differences in English hospital death rates using routinely collected data. *BMJ* 318: 1515-1520.

2 Jarman B, Pieter D, Veen AA, Kool RB, Aylin P, et al. (2010) The hospital standardised mortality ratio: a powerful tool for Dutch hospitals to assess their quality of care? *Qual Saf Health Care* 19: 9-13.

3 Wen E, Sandoval C, Zelmer J, Webster G (2008) Understanding and Using the Hospital Standardized Mortality Ratio in Canada: Challenges and Opportunities. *HealthcarePapers* 8(4): 26-36.

4 McKinley J, Gibson D, Ardal S (2008) Hospital Standardized Mortality Ratio: The Way Forward in Ontario. *HealthcarePapers* 8(4): 43-49.

5 Figler S. (2008) Data May Reveal Real Issues. *HealthcarePapers* 8(4): 54-56.

6 Zahn C, Baker M, MacNaughton J, Flemming C, Bell R (2008) Hospital Standardized Mortality Ratio Is a Useful Burning Platform. *HealthcarePapers* 8(4): 50-53.

7 Brien SE, Ghali WA (2008) CIHI's Hospital Standardized Mortality Ratio: Friend or Foe? *HealthcarePapers* 8(4): 57-61.

8 Brown C, Lilford R (2006) Cross sectional study of performance indicators for English Primary Care Trusts: testing construct validity and identifying explanatory variables. *BMC Health Serv Res* 6: 81.

9 Werner RM, Bradlow ET (2006) Relationship Between Medicare's Hospital Compare Performance Measures and Mortality Rates. *JAMA 296*: 2694-2702.

10 Pitches DW, Mohammed MA, Lilford RJ (2007) What is the empirical evidence that hospitals with higher-risk adjusted mortality rates provide poorer quality care? A systematic review of the literature. *BMC Health Serv Res* 7: 91.

11 Lilford RJ, Brown CA, Nicholl J (2007) Use of process measures to monitor the quality of clinical practice. *BMJ* 335: 648-650.

12 Mohammed MA, Deeks JJ, Girling A, Rudge G, Carmalt M, et al. (2009) Evidence of methodological bias in hospital standardized mortality ratios: retrospective database study of English hospitals. *BMJ* 338: b780.

13 Rothman KJ (1986) Modern Epidemiology. Boston: Little, Brown, and Company. 45-49 p.

14 Julious SA, Nicholl J, George S (2001) Why do we continue to use standardized mortality ratios for small area comparisons? *J Public Health Med* 23(1): 40-46.

15 Julious SA, George S (2007) Are hospital league tables calculated correctly? P*ublic Health* 121(12): 902-904.

16 Rixom A (2002) Performance league tables: use of indirect standardisation is inappropriate. *BMJ* 325: 177-178.

17 Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, et al. (2004) New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *J Clin Epidemiol*. 57: 1288-1294.

18 Charlson ME, Pompei P, Ales KL, MacKenzie CR (1987) A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 40: 373-383.

19 Heijink R, Koolman X, Pieter D, van der Veen A, Jarman B, et al. (2008) Measuring and explaining mortality in Dutch hospitals; the hospital standardized mortality rate between 2003 and 2005. *BMC Health Serv Res*. 8: 73.

APPENDIX S1

Suppose hospital A distinguishes two kinds of patients: patients admitted urgently and non-urgently admitted patients. Suppose that in year t hospital A has admitted 20% of the 5000 patients urgently. Let us now assume that urgently admitted patients have an expected mortality probability of 6% and non-urgently admitted patients have an expected mortality probability of 2% (see also table A.1). Furthermore, we assume that the *observed* mortality rates for these groups are 3% and 4% respectively.

Suppose that in year t+1 the quality of care of hospital A drops, expressed as increased observed mortality rates of 4% and 5% for urgently admitted patients and for non-urgently admitted patients respectively (see table A.1). Furthermore the number of urgently admitted patients and non-urgently admitted patients for that year changes to 2000 and 3000 respectively. This yields a HSMR of 128, which is lower than 136 in year *t* and suggests an increase in quality of care rather than a decrease. Hence, theoretically, a decreased performance of quality of care, reflected as an increase in observed mortality rates, can remain unnoticed due to differences in case-mix distribution.

Table A.1 Numerical example of HSMRs of hospital A in two consecutive years

	Hospital A (year t)		Hospital A (year t+1)	
	Urgent	Non-urgent	Urgent	Non-urgent
Expected mortality rate	6%	2%	6%	2%
Observed mortality rate	3%	4%	4%	5%
Case-mix	1000	4000	2000	3000
HSMR	136 1000 x 3% + 4000 x 4% 1000 x 6% + 4000 x 2%		128 2000 x 4% + 3000 x 5% 2000 x 6% + 3000 x 2%	

Despite a drop in quality of care, reflected as higher observed mortality rates, the HSMR improves in year t+1 because of a different casemix distribution.

APPENDIX S2

Table A.2 Scenario analysis HSMR

Original		Scenario 1			Scenario 5		
Ranking	HSMR	HSMR	New ranking	Rank difference	HSMR	New ranking	Rank difference
1	68 (*)	69 (*)	2	-1	69 (*)	1	0
2	69 (*)	68 (*)	1	1	73 (*)	2	0
3	79 (*)	83 (*)	5	-2	77 (*)	4	-1
4	79 (*)	79 (*)	3	1	76 (*)	3	1
5	82 (*)	82 (*)	4	1	82 (*)	5	0
6	85 (*)	84 (*)	6	0	86 (*)	7	-1
7	86 (*)	86 (*)	8	-1	85 (*)	6	1
8	87 (*)	86 (*)	7	1	91 (*)	16	-8
9	87 (*)	88 (*)	11	-2	90 (*)	15	-6
10	88 (*)	88 (*)	10	0	87 (*)	8	2
11	88 (*)	87 (*)	9	2	89 (*)	10	1
12	89 (*)	90 (*)	13	-1	88 (*)	9	3
13	90	90	12	1	89	11	2
14	90 (*)	91 (*)	15	-1	90 (*)	14	0
15	91 (*)	91 (*)	14	1	92	17	-2
16	91	91	17	-1	90	12	4
17	92 (*)	92 (*)	18	-1	92 (*)	18	-1
18	92	91	16	2	93	19	-1
19	92	92	19	0	95	23	-4
20	93	94	21	-1	94	21	-1
21	94	93	20	1	107	44	-23
22	95	95	25	-3	96	25	-3
23	95	96	26	-3	123 (**)	57	-34
24	96	94	23	1	96	26	-2
25	96	94	22	3	95	24	1
26	96	96	30	-4	94	22	4
27	97	96	29	-2	96	27	0
28	97	96	28	0	97	28	0
29	97	94	24	5	90 (*)	13	16
30	97	96	31	-1	98	30	0
31	97	96	27	4	101	35	-4
32	98	97	32	0	100	32	0

Original		Scenario 1			Scenario 5		
Ranking	HSMR	HSMR	New ranking	Rank differ-ence	HSMR	New ranking	Rank differ-ence
33	98	98	33	0	98	29	4
34	100	100	34	0	94	20	14
35	100	100	36	-1	100	33	2
36	101	102	41	-5	107	46	-10
37	101	100	35	2	101	34	3
38	101	101	38	0	101	36	2
39	101	101	37	2	99	31	8
40	102	103	42	-2	102	37	3
41	102	102	40	1	118 (**)	55	-14
42	102	102	39	3	104	39	3
43	103	103	43	0	104	40	3
44	103	104	46	-2	107	45	-1
45	104	103	44	1	103	38	7
46	105	104	45	1	104	41	5
47	105	105	48	-1	105	42	5
48	105	105	47	1	106	43	5
49	107	107	49	0	114 (**)	52	-3
50	108	108	51	-1	108 (**)	47	3
51	109	108	52	-1	113 (**)	51	0
52	109	109	53	-1	110	49	3
53	110	107	50	3	110	48	5
54	110 (**)	110 (**)	54	0	110 (**)	50	4
55	114 (**)	114 (**)	56	-1	129 (**)	60	-5
56	115	114	55	1	125 (**)	58	-2
57	120 (**)	121 (**)	58	-1	114 (**)	53	4
58	120 (**)	120 (**)	57	1	117 (**)	54	4
59	122 (**)	125 (**)	59	0	123 (**)	56	3
60	128 (**)	128 (**)	60	0	129 (**)	59	1
61	132 (**)	132 (**)	61	0	155 (**)	61	0

The HSMR of scenario 1 is computed based on the mean of the casemix distributions of the 'urgency of admission' variable of the 61 hospitals. The HSMR of scenario 5 is computed based on the mean of the casemix distributions of the 'Charlson Comorbidity index' variable of the 61 hospitals. (*) Significantly lower than 100, (**) significantly higher than 100.

CHAPTER 3

The impact of using Charlson versus Elixhauser comorbidity measures in hospital standardised mortality ratios

Pouw ME, Peelen LM, Moons KGM, Kalkman CJ

submitted

ABSTRACT

Objective

To compare the impact of the Charlson versus the Elixhauser comorbidity measure – as casemix adjustment tools – in calculating and comparing standardised mortality ratios of hospitals.

Design

Retrospective analysis of routinely collected hospital data comparing observed deaths with deaths predicted after casemix adjustment using either the Charlson comorbidity measure or the Elixhauser comorbidity measure.

Setting

86 Dutch hospitals.

Participants

2,387,352 discharges between 2007 and 2010.

Main outcome measures

Discriminative performance (c-statistic) and standardised mortality ratios based on (1) Charlson comorbidity risk adjustment models and on (2) Elixhauser comorbidity risk adjustment models.

Results

For 50 diagnoses, prediction models were estimated using either the Charlson comorbidity measure or the Elixhauser measure. The average difference in c-statistic between Charlson and Elixhauser models was 0.00. The maximum difference in c-statistic was 0.06 (0.68 vs. 0.62 for respectively Charlson and Elixhauser for diagnosis 'liver disease; alcohol related'). The maximum difference in which the c-statistic of the Elixhauser measure outperformed the c-statistic of the Charlson measure was 0.02. The confidence intervals of the c-statistic of all 50 Charlson models overlapped the confidence intervals of the c-statistic of the corresponding Elixhauser models. The HSMRs calculated with the Charlson and Elixhauser method did not differ significantly for all participating hospitals.

Conclusions

The discriminative performance of models estimated with the Charlson comorbidity measure and the Elixhauser comorbidity measure are comparable. Furthermore, the effects of the two measures on HSMRs of individual hospitals appear generally consistent. Because the Elixhauser measure is better suited when using administrative databases, as is the case in HSMRs, we recommend the Elixhauser measure in calculation of HSMRs.

INTRODUCTION

The hospital standardised mortality ratio (HSMR) is an important hospital outcome measure used to benchmark hospitals and to assess their quality of care.[1-4] Inappropriate calculation of the HSMR can have important consequences for hospitals in an environment of public reporting and pay-for-performance. If, however, differences in patient characteristics are appropriately taken into account, a meaningful comparison between hospitals' HSMRs can become more feasible. Current casemix adjustment methods are, nevertheless, still questioned and therefore the validity and ability of HSMRs to reflect quality of care is debated.[5-10]

Two commonly used and validated measures to express comorbidity were developed by Charlson et al. and Elixhauser et al.[11, 12] Both are lists of selected medical conditions aiming to capture the comorbidity load that exists next to the primary diagnosis. These lists are based on the International Classification of Diseases diagnosis codes used in administrative datasets.[13] Charlson et al. developed their measure for the purpose of predicting 1-year mortality among clinical patients with breast cancer.[11] Since then, numerous studies applied the Charlson measure to predict short-term outcomes such as in-hospital mortality and length of stay in various other clinical domains.[14-17]

The Elixhauser measure has been developed to predict length of stay, and in-hospital mortality using *administrative databases*. As administrative databases are criticised of being inaccurate and incomplete, Elixhauser et al. attempted to isolate complications from comorbidities by excluding conditions that could be the result from medical misadventures rather than pre-existing comorbidities.[12] Comorbidities were not taken into account if it was unclear whether they were already present at admission (true comorbidity) or whether they originated during hospitalisation (complication). Examples are urinary tract infection, pneumonia, respiratory failure, cardiac arrest, and cardiogenic shock. This represents a conservative approach as these conditions could be comorbidities, but given their nature, are often primary diagnoses or complications.

Previous studies have compared the predictive performance of these two measures in selected patient populations such as acute myocardial infarction (AMI), chronic obstructive pulmonary disease (COPD), acute cerebrovascular disease (CVA), and ICU admissions.[18-22] In general, these studies found that the Elixhauser measure outperformed the Charlson measure, particularly in terms of properly predicting mortality (discrimination, i.e. whether the model differentiates between subject who will die and those who survive).

The predictive performance of these measures, however, has never been compared in a hospital wide population. The aim of our study was twofold: (1) to assess the difference in discriminative performance between the Charlson measure and the Elixhauser measure when they are used for benchmarking purposes. (2) To assess to what extent these differences result in differences in HSMRs of individual hospitals.

METHODS

Data

The National Hospital Discharge Register (HDR) contains discharge data of Dutch hospitals and comprises patient characteristics such as age and gender, as well as medical variables such as date of admission, date of discharge, diagnoses, and comorbidities. The HDR follows the ICD9-CM (international classification of diseases, 9th revision, clinical modification) to register discharge diagnoses and comorbidities. Participation of hospitals in the HDR is voluntary. This study applied HDR data for hospital admissions during the period 2007 – 2010 from 86 out of 93 Dutch hospitals, as used in previous publications.[23, 24]

Charlson and Elixhauser comorbidities

The Charlson measure includes 17 comorbidities that are identified using ICD9-CM diagnosis codes. Typically, they are summed into a weighted overall score: the Charlson comorbidity index. For purpose of comparison with the Elixhauser measure, we assessed these 17 comorbidities individually rather than as a summarized weighted Charlson comorbidity score.

The Elixhauser measure includes 31 comorbidities that are identified using ICD9-CM diagnosis codes.[12] The Elixhauser measure does not have a weighted scoring system and permits each comorbidity to be assessed individually in models. The comorbidities that are scored within the Charlson and Elixhauser measures are listed in table 3.1.

Risk adjusted mortality models

The Dutch HSMR method is based on the HDR database and developed by Statistics Netherlands.[25] It uses 50 risk-adjustment models – each for one of 50 selected diagnostic groups – to estimate the risk of mortality for patients belonging to the specific diagnostic group. These 50 diagnostic groups account for approximately 80% of all in-hospital deaths in the Netherlands.[25]

The models are logistic regression models with mortality as the dependent variable and age, sex, socio-economic status, severity of main diagnosis, urgency of admission, source of admission, year and month of admission, and a comorbidity measure as predictor variables.

For the current study we developed two sets of models, one using the Charlson measure (Charlson model), and one using the Elixhauser measure (Elixhauser model).

Discriminative performance of risk adjusted mortality models

To compare the predictive performances of the Charlson models with the Elixhauser models, we calculated the c-statistic (with its 95% confidence interval). The c-statistic is a measure of a model's ability to discriminate those who die of those who do not die. The potential values for the c-statistic range from 0.5 (no greater predictive power than chance) to 1.0 (perfect prediction).

Comparison of SMRs

The SMR of a diagnostic group of a hospital is the ratio of the observed number of deaths and the expected number of deaths as calculated with the prediction model. The sum of the observed number of mortalities of all 50 diagnostic groups divided by the sum of all expected number of mortalities times 100 gives the hospital wide SMR (HSMR). A SMR greater than 100 indicates higher observed mortality than expected.

Plots were made to evaluate the magnitude and direction of change in SMR when substituting SMR based on the Charlson model for a SMR with the Elixhauser model. Furthermore, hospitals were classified into three groups as follows: if the 95% confidence interval of the SMR included the reference value of 100, the hospital was categorized into the group 'as expected'. A hospital was regarded 'better than expected' or 'worse than expected' if the

95% confidence interval of the SMR was respectively below 100 or above 100. We analysed how many hospitals would be classified differently with the Charlson model and the corresponding Elixhauser model.

RESULTS

Data

This study included 2,387,352 discharges of 86 hospitals. Patient characteristics including comorbidities are presented in table 3.1. The average age was 63.6 years, 52.7% of the discharges were male and 58.8% of the discharges were admitted urgently. Some comorbidities are named similar in the Charlson measure and the Elixhauser measure but differ in the included ICD-9 diagnoses. Therefore the number of comorbidities can differ between the Charlson and the Elixhauser measure, e.g. the comorbidity 'congestive heart failure' has been recorded in 45,788 (1.9%) admissions with the Charlson definition and in 32,143 (1.4%) admissions with the Elixhauser definition. The most often registered Charlson comorbidity is 'pulmonary disease' (3.4%) and the most often registered Elixhauser comorbidity is 'hypertension uncomplicated' (3.5%).

Table 3.1 Baseline characteristics and Charlson and Elixhauser comorbidities

Baseline characteristics	N = 2,387,352 (%)
Hospitals	86
Age	63.56
Male Sex	1,258,134 (52.7)
Urgent admissions	1,403,763 (58.8)
Mortality	113,305 (4.7)
Charlson Comorbidities	**Cases with comorbidity**
1. Acute myocardial infarction	68,047 (2.9)
2. Congestive heart failure	45,788 (1.9)
3. Peripheral vascular disease	29,993 (1.3)
4. Cerebral vascular accident	25,732 (1.1)
5. Dementia	12,247 (0.5)
6. Pulmonary disease	81,076 (3.4)
7. Connective tissue disorder	7,922 (0.3)
8. Peptic ulcer	3,139 (0.1)
9. Liver disease	5,126 (0.2)
10. Severe liver disease	79,175 (3.3)
11. Diabetes	6,583 (0.3)
12. Diabetes complications	7,587 (0.3)
13. Paraplegia	35,788 (1.5)
14. Renal disease	65,332 (2.7)
15. Cancer	1,111 (0.0)
16. Metastatic cancer	142,636 (6.0)
17. HIV	2,682 (0.1)

Elixhauser comorbidities	Cases with comorbidity
1. Congestive heart failure	32,143 (1.4)
2. Cardiac arrhythmias	28,917 (1.2)
3. Valvular disease	59,531 (2.5)
4. Pulmonary circulation disorders	7,193 (0.3)
5. Peripheral vascular disorders	31,732 (1.3)
6. Hypertension uncomplicated	83,868 (3.5)
7. Hypertension complicated	662 (0.0)
8. Paralysis	3,064 (0.1)
9. Other neurological disorders	18,831 (0.8)
10. Chronic pulmonary disease	18,710 (0.8)
11. Diabetes, complicated	15,701 (0.7)
12. Diabetes, uncomplicated	3,996 (0.2)
13. Hypothyroidism	761 (0.0)
14. Renal failure	9,972 (0.4)
15. Liver disease	2,631 (0.1)
16. Peptic ulcer disease (excluding bleeding)	356 (0.0)
17. AIDS/HIV	243 (0.0)
18. Lymphoma	2,822 (0.1)
19. Metastatic cancer	43,643 (1.8)
20. Solid tumour without metastasis	40,529 (1.7)
21. Rheumatoid arthritis / collagen vascular disease	1,894 (0.1)
22. Coagulopathy	1,349 (0.1)
23. Obesity	2,344 (0.1)
24. Weight loss	81 (0.0)
25. Fluid and electrolyte disorder	7,396 (0.3)
26. Blood loss anaemia	793 (0.0)
27. Deficiency anaemia	8,838 (0.4)
28. Alcohol abuse	3,881 (0.2)
29. Drug abuse	369 (0.0)
30. Psychoses	1,348 (0.1)
31. Depression	549 (0.0)

Performance of risk adjusted mortality models

Table 3.2 describes the c-statistics for the 50 models using Charlson and Elixhauser measures respectively. For 33 diagnoses the c-statistic of the Charlson model was higher than of the Elixhauser model. For eight Charlson models the c-statistic was more than 0.01 points higher than for the corresponding Elixhauser models. For three Elixhauser models the c-statistic was more than 0.01 points higher than for the corresponding Charlson models. The maximum difference in c-statistic was 0.06 (0.68 vs 0.62 for respectively Charlson and Elixhauser for the diagnostic group 'liver disease; alcohol related'). The maximum difference in models where the Elixhauser measure outperformed the Charlson measure was 0.02 (diagnostic group 'coronary atherosclerosis and other heart disease').

Table 3.2 c-statistics (with .95 confidence interval) of the 50 diagnostic groups using Charlson models and Elixhauser models.

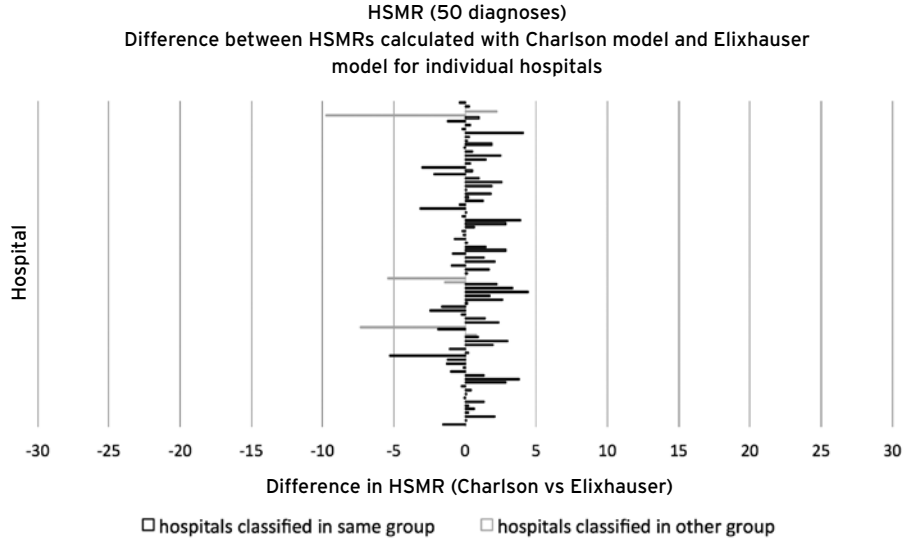
Description CCS diagnosis groups	N	Charlson measure		Elixhauser measure	
		c-stat	(C.I)	c-stat	(C.I)
Septicaemia (except in labour)	15447	**0.75**	(0.74 - 0.76)	**0.74**	(0.73 - 0.75)
Cancer of oesophagus	10591	**0.74**	(0.72 - 0.76)	**0.74**	(0.72 - 0.76)
Cancer of stomach	14562	**0.78**	(0.77 - 0.80)	**0.78**	(0.76 - 0.79)
Cancer of colon	40196	**0.78**	(0.77 - 0.79)	**0.78**	(0.77 - 0.79)
Cancer of rectum and anus	21204	**0.79**	(0.77 - 0.81)	**0.79**	(0.77 - 0.80)
Cancer of pancreas	10772	**0.71**	(0.69 - 0.73)	**0.71**	(0.69 - 0.72)
Cancer of bronchus; lung	72235	**0.81**	(0.81 - 0.82)	**0.81**	(0.80 - 0.81)
Cancer of breast	57519	**0.92**	(0.91 - 0.93)	**0.92**	(0.91 - 0.94)
Cancer of prostate	23077	**0.89**	(0.88 - 0.90)	**0.89**	(0.87 - 0.90)
Cancer of bladder	41505	**0.87**	(0.86 - 0.89)	**0.87**	(0.86 - 0.89)
Non-Hodgkin's lymphoma	20103	**0.81**	(0.80 - 0.82)	**0.80**	(0.79 - 0.82)
Leukaemia	19217	**0.83**	(0.82 - 0.84)	**0.83**	(0.82 - 0.84)
Secondary malignancies	67708	**0.77**	(0.76 - 0.77)	**0.76**	(0.76 - 0.77)
Neoplasms of unspecified nature	20892	**0.78**	(0.76 - 0.80)	**0.77**	(0.75 - 0.80)
Diabetes mellitus with complications	32490	**0.85**	(0.84 - 0.86)	**0.84**	(0.83 - 0.85)
Fluid and electrolyte disorders	26918	**0.78**	(0.77 - 0.79)	**0.77**	(0.76 - 0.79)
Deficiency and other anaemia	45375	**0.75**	(0.73 - 0.77)	**0.75**	(0.73 - 0.77)
Coma; stupor; and brain damage	3982	**0.69**	(0.67 - 0.71)	**0.67**	(0.65 - 0.69)
Heart valve disorders	33771	**0.77**	(0.76 - 0.79)	**0.78**	(0.77 - 0.79)
Acute myocardial infarction	87839	**0.74**	(0.74 - 0.75)	**0.76**	(0.75 - 0.76)
Coronary atherosclerosis and other heart disease	228049	**0.79**	(0.78 - 0.80)	**0.80**	(0.79 - 0.81)

Description CCS diagnosis groups	N	Charlson measure		Elixhauser measure	
		c-stat	(C.I)	c-stat	(C.I)
Pulmonary heart disease	26059	**0.77**	(0.76 - 0.78)	**0.77**	(0.76 - 0.78)
Cardiac dysrhythmias	180943	**0.83**	(0.82 - 0.84)	**0.83**	(0.82 - 0.84)
Cardiac arrest and ventricular fibrillation	8251	**0.77**	(0.76 - 0.78)	**0.78**	(0.77 - 0.79)
Congestive heart failure; non hypertensive	99702	**0.65**	(0.65 - 0.66)	**0.66**	(0.65 - 0.66)
Acute cerebrovascular disease	95050	**0.65**	(0.65 - 0.66)	**0.65**	(0.64 - 0.65)
Peripheral and visceral atherosclerosis	40468	**0.87**	(0.86 - 0.88)	**0.87**	(0.86 - 0.88)
Aortic; peripheral; and visceral artery aneurysms	26278	**0.83**	(0.83 - 0.84)	**0.83**	(0.83 - 0.84)
Aortic and peripheral arterial embolism or thrombosis	28566	**0.88**	(0.87 - 0.89)	**0.88**	(0.87 - 0.89)
Other circulatory disease	21763	**0.80**	(0.79 - 0.82)	**0.80**	(0.78 - 0.81)
Pneumonia	118900	**0.77**	(0.77 - 0.78)	**0.76**	(0.75 - 0.77)
Chronic obstructive pulmonary disease	76089	**0.70**	(0.69 - 0.71)	**0.70**	(0.69 - 0.71)
Aspiration pneumonitis; food/vomitus	4517	**0.68**	(0.66 - 0.70)	**0.66**	(0.65 - 0.68)
Pleurisy; pneumothorax; pulmonary collapse	22679	**0.82**	(0.81 - 0.83)	**0.82**	(0.81 - 0.83)
Other lower respiratory disease	97751	**0.84**	(0.83 - 0.84)	**0.84**	(0.83 - 0.84)
Intestinal obstruction without hernia	32209	**0.83**	(0.82 - 0.83)	**0.83**	(0.83 - 0.84)
Diverticulosis and diverticulitis	32655	**0.84**	(0.83 - 0.86)	**0.84**	(0.83 - 0.86)
Biliary tract disease	122761	**0.90**	(0.90 - 0.92)	**0.90**	(0.90 - 0.92)
Liver disease; alcohol-related*	4848	**0.68**	(0.65 - 0.70)	**0.62**	(0.60 - 0.65)
Other liver diseases	16365	**0.72**	(0.71 - 0.74)	**0.70**	(0.69 - 0.71)
Gastrointestinal haemorrhage	33655	**0.78**	(0.77 - 0.79)	**0.76**	(0.74 - 0.77)
Other gastrointestinal disorders	45740	**0.90**	(0.89 - 0.91)	**0.90**	(0.89 - 0.91)
Acute and unspecified renal failure	9762	**0.75**	(0.73 - 0.76)	**0.74**	(0.72 - 0.76)
Chronic renal failure	18678	**0.85**	(0.84 - 0.87)	**0.85**	(0.84 - 0.86)
Urinary tract infections	62258	**0.83**	(0.82 - 0.84)	**0.82**	(0.81 - 0.82)
Fracture of neck of femur (hip)	66409	**0.77**	(0.76 - 0.78)	**0.76**	(0.76 - 0.77)
Intracranial injury	56982	**0.91**	(0.90 - 0.92)	**0.91**	(0.90 - 0.92)
Complication of device; implant or graft	74620	**0.81**	(0.80 - 0.82)	**0.81**	(0.80 - 0.82)
Complications of surgical procedures or medical care	67030	**0.83**	(0.82 - 0.84)	**0.82**	(0.81 - 0.83)
Shock	2912	**0.74**	(0.72 - 0.75)	**0.71**	(0.70 - 0.73)

Comparison of SMRs

Figure 3.1 shows the difference in HSMR for individual hospitals when using the Charlson model or the Elixhauser model. The HSMR calculated with the Charlson and Elixhauser model did not differ significantly for all hospitals. For the 86 hospitals, the average difference between HSMRs calculated with the Charlson and Elixhauser models was 1.6 points. Four hospitals classified 'better than expected' with the Charlson model were classified 'as expected' with the Elixhauser model. One hospital classified 'as expected' with the Charlson model was classified 'worse than expected' with the Elixhauser model and one hospital moved into the opposite direction (from 'worse than expected' to 'as expected').

Figure 3.1 Difference in HSMRs calculated with the Charlson model and the Elixhauser model



Figures 3.2, 3.3, 3.4 and 3.5 show the difference in SMR for individual hospitals when using the two different measures for the diagnoses 'acute myocardial infarction', 'congestive heart failure', 'cerebrovascular accident', and 'COPD'. Although there is considerable variation in the SMRs for these diagnoses, these differences were not statistically significant for any of the individual hospitals. For the diagnoses 'acute myocardial infarction', 'congestive heart failure', 'cerebrovascular accident', and 'COPD' respectively 3, 13, 7, and 6 hospitals would be classified differently when using the Elixhauser model instead of the Charlson model.

Figure 3.2 Difference in SMRs calculated with the Charlson model and the Elixhauser model

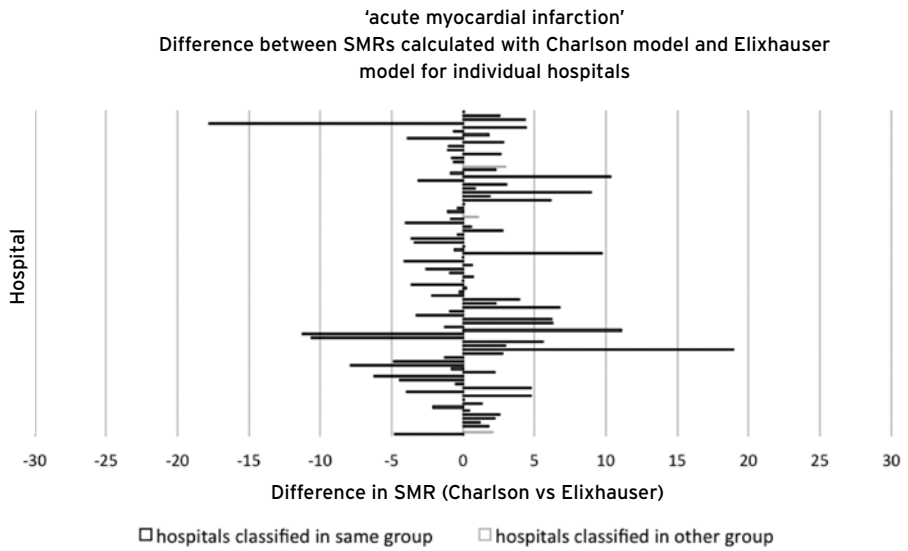


Figure 3.3 Difference in SMRs calculated with the Charlson model and the Elixhauser model

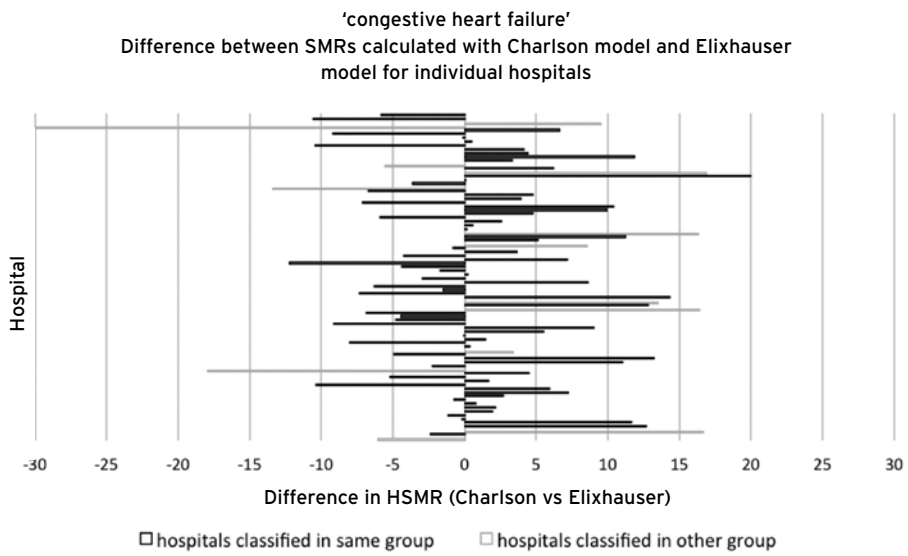


Figure 3.4 Difference in SMRs calculated with the Charlson model and the Elixhauser model

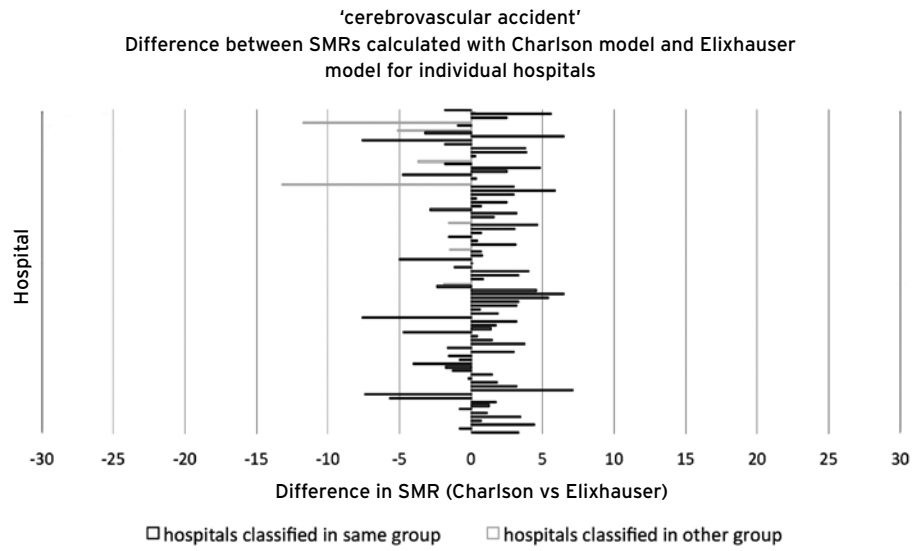
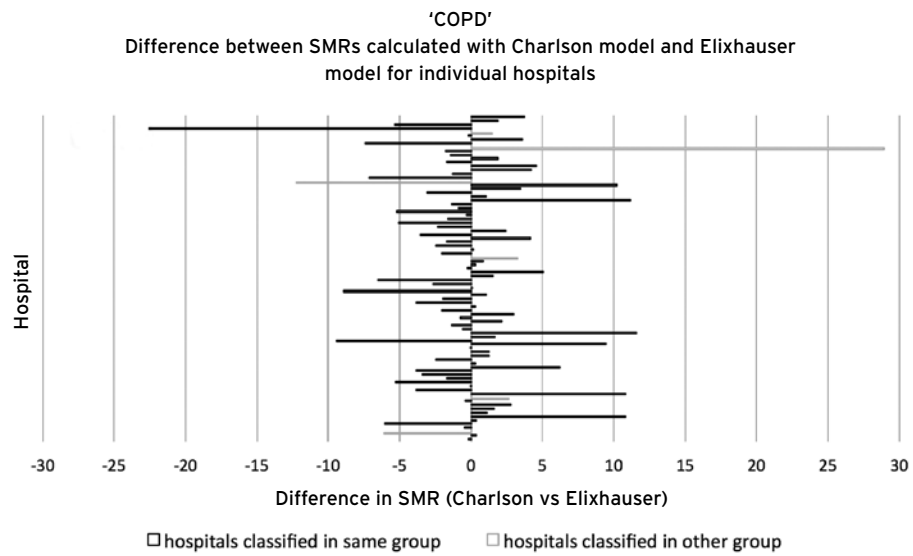


Figure 3.5 Difference in SMRs calculated with the Charlson model and the Elixhauser model



DISCUSSION

Our study suggests that – using administrative data – the difference in discriminative performance of models estimated with the Charlson comorbidity measure and the Elixhauser comorbidity measure is marginal. Furthermore, the effects of the two measures on (H)SMRs of individual hospitals appear generally consistent.

In general, our results agree with previous work, which indicate that the discriminative performance of the Elixhauser measure is at least equal to the Charlson measure in predicting mortality.[18, 20, 21] However, we did not find strong evidence that the Elixhauser measure *outperforms* the Charlson measure as was claimed in some studies. For example, Southern et al. concluded that – based on Canadian administrative data – the Elixhauser measure outperforms the Charlson measure in predicting mortality on patients with acute myocardial infarction (AMI) (c-statistic of 0.79 vs 0.70).[20] Chu et al also compared the Charlson measure with the Elixhauser measure using a Taiwanese National Health Insurance administrative database to predict the in-hospital mortality for patients admitted with AMI and chronic obstructive pulmonary disease (COPD). They found that the c-statistic was higher with the Elixhauser measure than with the Charlson measure for both AMI and COPD (respectively 0.74 vs 0.71 for both diagnoses).[21] Stukenborg et al. also analysed the performance of the Charlson and Elixhauser measure for AMI and COPD when estimating hospital mortality. In addition they also examined the diagnoses 'congestive heart failure' and 'acute cerebrovascular disease'. They concluded that for all four diagnoses the discriminative performance of Elixhauser outperformed that of Charlson.[18] When focusing on these four diagnoses, in our analysis, the c-statistic of the model predicting hospital mortality in AMI patients was only 0.02 higher with the Elixhauser measure than with the Charlson measure (respectively 0.76 (CI: 0.75-0.76) and 0.74 (CI: 0.74-.075)). For the diagnosis COPD, 'acute cerebrovascular disease', and 'congestive heart failure' we also found relatively small differences in c-statistics. Furthermore, the estimations of Charlson and Elixhauser models for these four diagnoses did not result in significantly different SMRs for all hospitals.

Instead of focusing on a single diagnosis or a small group of diagnoses like in previous studies, we observed that the Elixhauser model did not consistently outperform the models with the Charlson measure when investigating 50 different diagnoses. In that respect, the present study is, to our knowledge, the first to compare the Elixhauser and Charlson measures on a hospital wide population.

Furthermore, we examined the effect of these measures on (H)SMRs of hospitals. Therefore, our study gives a broader and deeper insight in (1) the difference in model performance with the Charlson and Elixhauser measures and (2) the implications of these differences for individual hospitals when using SMRs.

Limitations of our study include the fact that the outcome is in-hospital mortality. Previous studies have shown why mortality at fixed time intervals is a more appropriate measure in outcomes evaluation.[23, 26, 27] For the purpose of our study, we have no reason to believe this has affected our results, as both the Charlson and Elixhauser models were fitted on this outcome. Second, not all Dutch hospitals could be examined as participation in the HDR-database was on a voluntary basis. SMRs of these hospitals could potentially be assessed differently with the Charlson and Elixhauser models. However, again we have no reason to believe that inclusion of data of non-participating hospitals would have changed our results as both type of models were fitted on the same dataset.

Theoretically, the performance of Elixhauser models in terms of discrimination might be worse than Charlson models because of the conservative approach of the Elixhauser measure: comorbidities that cannot be distinguished from complications are not included, although they may have significant value in estimating mortality (e.g. cardiogenic shock and pneumonia). Nevertheless, in our study we see only marginal differences in c-statistics between the Charlson measure and the Elixhauser measure. Moreover, in our current analysis the performance of the Elixhauser models may be even underestimated, as administrators of hospitals currently pay specific attention on comorbidities included in the Charlson measure. Comorbidities included in the Elixhauser but not in the Charlson measure might hence be under-registrated. For example, for obesity, a comorbidity that is part of Elixhauser but not of Charlson, the prevalence in the Dutch population was over 10% in the years 2005-2010.[28] However, in our database only 2,344 (0.1%) admissions were coded with obesity.

Conclusion

In conclusion, models estimated with the Charlson measure and the Elixhauser measure perform practically similarly in terms of discrimination. Furthermore, because the Elixhauser measure tries to differentiate complications from comorbidities, expected mortality estimated with the Elixhauser model would theoretically be more fair. For these reasons, we recommend to use the Elixhauser measure when estimating casemix correction models to calculate SMRs.

REFERENCE LIST

1 Jarman B, Galt S, Alves B, Hider A, Dolan S, Cook A, et al. Explaining differences in English hospital death rates using routinely collected data. *BMJ* 1999; 318:1515-20.

2 Jarman B, Pieter D, et al. The hospital standardised mortality ratio: a powerful tool for Dutch hospitals to assess their quality of care? *Qual Saf Health Care* 2010; 19:9-13.

3 Ben-Tovim D, Woodman R, Harrison J, et al. Measuring and reporting mortality in hospital patients. Canberra: Australian Institute of Health and Welfare, 2009. (AIHW Cat. No. HSE 69.)

4 Canadian Institute for Health Information. HSMR Technical notes, updated September 2012. Ottawa 2012, http://www.cihi.ca/CIHI-ext-portal/pdf/internet/HSMR_TECH_NOTES_201202_en.pdf

5 Tu Y-K, Gilthorpe MS. The most dangerous hospital or the most dangerous equation? *BMC Health Serv Res* 2007; 7:185-9.

6 Powell H, Li LL, Heller RF. Accuracy of administrative data to assess comorbidity in patients with heart disease: an Australian perspective. *J Clin Epidemiol* 2001; 54:687-93.

7 Van Gestel YRBM, Lemmens EPP, Lingsma HF, de Hingh IHJT, Rutten HJT, Coebergh JW. The Hospital Standardised Mortality Ratio Fallacy, a narrative review. *Med Care* 2012; 50:662-7

8 Van den Bosch WF, Kelder JC, Wagner C. Predicting hospital mortality among frequently readmitted patients: HSMR biased by readmission. *BMC Health Serv Res* 2011; 11:57.

9 Mohammed MA, Deeks JJ, Girling A, et al. Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals. *BMJ* 2009; 338:b780.

10 Pouw ME, Peelen LM, Lingsma HF, Pieter D, Steyerberg E, Kalkman CJ, Moons KGM. Hospital standardized mortality ratio: consequences of adjusting hospital mortality with indirect standardization. *PLoS One* 2013; 8(4): e59160

11 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987; 40:373–83.

12 Elixhauser A, Steiner C, Harris R, Coffey R. Comorbidity measures for use with administrative data. *Med Care* 1998; 36:8–27.

13 World Health Organization. http://www.who.int/classifications/icd/en/

14 D'Hoore W, Bouckaert A, Tilquin C. Practical considerations on the use of the Charlson Comorbidity Index with administrative data bases. *J Clin Epidemiol* 1996; 49:1429

15 D'Hoore W, Sicotte C, Tilquin C. Risk adjustment in outcome assessment: The Charlson Comorbidity Index. *Methods Inf Med* 1993; 32:382

16 Hartz AJ, Kuhn EM, Kayser KL, Pryor DP, Green R, Rimm AA. Assessing providers of coronary revasularization: A method for peer review organizations. *Am J Public Health Res* 1992; 82:1631

17 Roos LL, Sharp SM, Cohen MM, Wajda A. Risk adjustment in claims-based research: The search for efficient approaches. *J Clin Epidemiol* 1989; 42:1193

18 Stukenborg GJ, Wagner DP, Connors AF. Comparison of the Performance of two Comorbidity Measures, with and without Information from Prior Hospitalizations, *Med Care* 2001; 39:727-39

19 Bottle A, Aylin P. Comorbidity scores for administrative data benefited from adaption to local coding and diagnostic practices, *J Clin Epidemiol* 2011; 64:1426-33

20 Southern DA, Quan H, Ghali, WA. Comparison of the Elixhauser and Charlson/Deyo Methods of Comorbidity Measurement in Administrative Data, *Med Care* 2004; 42: 355-60)

21 Chu YT, Yung Ng Y, Wu SC. Comparison of different comorbidity measures for use with administrative data in predicting short-and long-term mortality, *BMC Health Serv Res* 2010; 10:140

22 Sharabiani MTA, Bottle A, Aylin P. Systematic Review of Comorbidity Indices for Administrative Data, *Med Care* 2012; 50(12): 1109-18

23 Pouw ME, Peelen LM, Moons KG, et al. Including post-discharge mortality in calculation of hospital standardised mortality ratios: retrospective analysis of hospital episode statistics. *BMJ* 2013;347:f5913

24 Siregar S, Pouw ME, Moons KGM, et al. The Dutch Hospital Standardised Mortality Ratio (HSMR) method and cardiac surgery: benchmarking in a national cohort using hospital administration data versus a clinical database. *Heart* 2014; 100(9):702-710. doi:10.1136/heartjnl-2013-304645.

25 Israels A, van der Laan J, de Bruin A, Ploemacher J, Verweij G.HSMR 2010, methodological report. Statistical Methods (201122). http://www.cbs.nl/en-GB/menu/themas/gezondheid-welzijn/publicaties/artikelen/archief/2011/2011-hsmr2010-report.htm

26 Vasilevskis EE, Kuzniewicz MW, Dean ML, Clay T, Vittinghoff E, Rennie DJ, Dudley RA. Relationship between discharge practices and intensive care unit in-hospital mortality performance: evidence of a discharge bias. *Med Care* 2009; 47(7): 803-12.

27 Drye EE, Normand SL, Wang Y, Ross JS, Schreiner GC, Han L, Rapp M, Krumholz HM. Comparison of hospital risk-standardized mortality rates calculated by using in-hospital and 30-day models: an observational study with implications for hospital profiling. *Ann Intern Med* 2012; 156:19-26

28 http://www.cbs.nl/en-GB/menu/themas/gezondheid-welzijn/publicaties/artikelen/archief/2014/2014-3939-wm.htm

CHAPTER 4

Including post-discharge mortality in the calculation of hospital standardised mortality ratios: a retrospective analysis of hospital episode statistics

Pouw ME, Peelen LM, Moons KGM, Kalkman CJ, Lingsma HF

BMJ 2013; 347: f5913

ABSTRACT

Objectives

To assess the consequences of applying different mortality timeframes on standardised mortality ratios of individual hospitals. A secondary objective was to evaluate the association between in-hospital standardised mortality ratios and the early post-discharge mortality rate, length of hospital stay, and transfer rate.

Design

Retrospective analysis of routinely collected hospital data to compare observed deaths in 50 diagnostic categories with deaths predicted by a casemix adjustment method.

Setting

60 Dutch hospitals in the period 2008 – 2010.

Participants

1,228,815 patient discharges in the period 2008 – 2010.

Main outcome measures

In-hospital standardised mortality ratio, 30 days post-admission standardised mortality ratio, and 30 days post-discharge standardised mortality ratio.

Results

Compared with in-hospital standardised mortality ratio, 33% of the hospitals were categorised differently with 30 days post-admission standardised mortality ratio and 21% were categorised differently with 30 days post-discharge standardised mortality ratio. There was a positive association (Pearson correlation coefficient = 0.33, P-value = 0.01) between in-hospital standardised mortality ratio and length of hospital stay, and an inverse association between in-hospital standardised mortality ratio and early post-discharge mortality (Pearson correlation coefficient = – 0.37, P-value = 0.004).

Conclusions

Applying different mortality timeframes resulted in differences in standardised mortality ratios and in differences in judgment regarding the performance of individual hospitals. Furthermore, associations between in-hospital standardised mortality rates, length of stay, and early post-discharge mortality rates were found. Combining these findings suggests that standardised mortality ratios based on in-hospital mortality are subject to so-called 'discharge bias'. Hence, early post-discharge mortality should be included in the calculation of standardised mortality ratios.

INTRODUCTION

The last decades, quality of care in hospitals has been subject to growing attention from physicians and regulators. In various countries standardised mortality rates (SMRs) are used in an attempt to judge the quality of hospital care.[1-4] However, several authors raised concerns that differences in SMRs may not reflect differences in delivered quality of care.[5,6] Reasons that have been put forward include the quality of the data used, the limitations of case mix adjustment, and several methodological issues.[7-12] Another limitation of mortality rate as a quality measure is the current focus on in-hospital mortality, i.e., deaths that occur during hospitalisation. Analyses based only on in-hospital deaths are potentially biased by differences in hospital discharge practices. For example, hospitals that transfer high-risk patients to other more specialized hospitals may have lower than expected mortality, because some of their patients die elsewhere. Furthermore, the average length of hospital stay (LoS) has decreased significantly the last decades and may therefore shift mortality away from the hospital to post-discharge destinations.[13,14] A recent study of Yu et al. (2011) revealed that for certain surgical procedures approximately one-fourth of postoperative deaths occurred after discharge and that 12% took place just one day after hospital discharge.[15] Metersky et al. (2012) concluded that approximately 50% of elderly patients, who died from pneumonia within 30 days of admission, did not die in hospital, but after discharge.[16] We will refer to this phenomenon as 'early post-discharge mortality'.

Differences in discharge practices or LoS between hospitals thus may affect their in-hospital mortality rate. Such biases arising from differences in discharge practices could have important consequences for hospitals in an environment of public reporting and pay-for-performance. When the timeframe to observe death is fixed or is prolonged to the post discharge period, these 'discharge' biases may be countered. For example, the UK used to report SMRs based on in-hospital mortality, but recently prolonged the timeframe to '30 days post-discharge'.[17] A commonly used alternative timeframe is the '30 days post-admission' timeframe that covers the fixed period from admission to 30 days post-admission.[15,16,18] In this study, we will explore both timeframes.

The aim of our study is to assess the effect of different mortality timeframes on SMRs of individual hospitals and judgment of their performance. A secondary objective was to investigate the relation between in-hospital SMR and early post-discharge mortality, LoS, and transfer rates. We used data from more than two million Dutch hospital discharges to explore the differences between hospital standardised mortality rates using either in-hospital mortality, 30 days post-admission mortality, and 30 days post-discharge mortality.

METHODS

Data

Dutch Hospital Data (DHD), the holder of the National Hospital Discharge Register (HDR), gave permission to use their database to perform this study. The HDR contains discharge data of general and academic Dutch hospitals and comprises patient characteristics such as age and gender as well as medical variables such as date of admission, date of discharge, diagnoses, and comorbidities. The HDR follows the ICD9-CM (international classification of diseases, 9th revision, clinical modification) to register discharge diagnoses. Participation of hospitals in the HDR is voluntary. In the period 2007- 2010, the total number of hospitals in the Netherlands was 100, of which 84 hospitals participated in the HDR and contributed to this study.

To obtain information on deaths that occurred after hospital discharge, records from the HDR were linked to the Dutch population register by Statistics Netherlands (www.CBS.nl). The population register contains personal details such as date of birth, date of death (if applicable), gender and address of all residents in The Netherlands. Because the HDR is pseudonomised, only date of birth, sex and truncated postal code (4 digits) are available for linkage with the population register. Statistics Netherlands regularly evaluates the linkage of the Dutch HDR with the population register and concludes that it is of good quality and forms an adequate basis for statistical analyses.[19] The combined dataset was used to compute the time to death (subtracting the date of admission from the date of mortality on the death certificate) and to fit the statistical models.

Risk adjusted mortality models

Statistics Netherlands calculates the Dutch hospital standardised mortality ratios each year as follows[20]: only in-patient records with a primary diagnosis belonging to one of 50 selected diagnostic groups (coded using Clinical Classification System, CCS[21]) were selected. These 50 diagnostic groups account for approximately 80% of all in-hospital deaths in the Netherlands. For each of the 50 selected diagnostic groups a prediction model is estimated to calculate the expected mortality probability of an admission. The models are logistic regression models with mortality as the dependent variable and age, sex, socio-economic status, severity of main diagnosis, urgency of admission, comorbidities,

source of admission, and month of admission as predictor variables. First, regression models are estimated using all predictors. Subsequently, reduced models are estimated, dropping non-significant variables using a backward stepwise elimination procedure.[22] The SMR of a diagnostic group of a hospital is the ratio of the observed number of deaths and the expected number of deaths as calculated based on the regression model. The sum of the observed mortalities of all 50 diagnostic groups divided by the sum of all expected mortalities times 100 gives the hospital wide SMR (HSMR). A SMR greater than 100 indicates higher mortality than expected.

To calculate SMRs with different timeframes, we recalibrated the 50 prediction models by redefining 'mortality' according to the used timeframe. In this recalibration procedure, we used the same variables as independent predictors and re-estimated the coefficients in the regression model. We chose not to include additional (post-discharge) variables in the post-discharge prediction models, because then it would be difficult to determine whether differences in SMRs are due to the different timeframe used or to the introduction of a new variable. All regression models were estimated in R version 2.15.

Comparison of hospital SMRs based on different mortality timeframes

We first examined to what extent SMRs depend on the mortality timeframe definition. Histograms and scatterplots were made to evaluate the magnitude and direction of performance change when substituting in-hospital SMR for a SMR with another timeframe. In addition, hospitals were classified into three groups based on the 95% confidence interval of their SMRs. If the 95% confidence interval of the SMR included the reference value of 100, the hospital was categorized into the group 'as expected'. A hospital was regarded 'better than expected' or 'worse than expected' if the confidence interval of the SMR was respectively below 100 or above 100. We analysed how many hospitals would be categorized differently when a different timeframe was used.

Effect of discharge patterns on in-hospital SMR

We examined the following variables for their association with in-hospital SMR: '*early* post-discharge' mortality rate (defined as mortality between discharge and 30 days post-admission divided by the number of alive discharges), average LoS, and average transfer rate to other medical facilities such as hospi-

tals, nursing homes and other medical institutions (excluding elderly homes). For all of these variables we evaluated the association with in-hospital SMR using the Pearson Correlation coefficient. All analyses were performed in SPSS 16.0.2.

RESULTS

Data and risk adjusted mortality models

The dataset contained 2,387,604 discharges of which 2,149,958 (90%) could be uniquely paired with the population register. Patient characteristics of the linkable discharges and non-linkable discharges are shown in table 4.1. In summary, the non-linkable discharges are on average younger, more often male, more often admitted with urgency, and have a lower in-hospital mortality rate.

Table 4.1 Baseline characteristics. Values are numbers (percentages) unless stated otherwise

Characteristics	Included admissions (n=2 149 958)	Excluded admissions (n=237 646)
Average age (years)	64.1	58.9
Male sex	1 013 519 (47.1)	114 841(48.3)
Urgent admission	1 260 927 (58.7)	142 895 (60.1)
Average length of stay (days)	7.5	8.0
In-hospital death	104 337 (4.9)	8987 (3.8)

Admissions between 2007 and 2010 were not included if no unique link was possible between Hospital Discharge Register database and population register.

We used HDR data of all patients discharged in the period 2007-2010 that could be uniquely paired with the population register to estimate the coefficients of the 50 prediction models for in-hospital mortality, 30 days post-admission mortality, and 30 days post-discharge mortality. The estimated models are available on request.

The Dutch hospitals are categorized in 8 academic hospitals, 84 general hospitals, and 8 specialized hospitals such as eye hospitals and epilepsy clinics. For all 8 academic hospitals and 52 general hospitals we analysed the 3-year SMR (2008 – 2010). Thirty-two general hospitals were excluded from analysis for one of three reasons: (1) no or insufficient participation in the HDR (e.g. start of participation after 2009), (2) inadequate data (e.g. no registration of comorbidities), (3) no permission to publish. The specialized hospitals were excluded from analysis because of their unique patient profiles[20]. We used the 3-year SMR because this is common practice in The Netherlands[20].

In 2008-2010 the 60 included hospitals discharged 1,228,815 patients. Across these hospitals mean in-hospital mortality rate for the 50 diagnostic groups was 4.9% (SD: 0.7%). Average LoS in 2008-2010 was 7.2 days (SD: 0.7 days). For the 60 hospitals 1,199,889 patients, 97.8% (SD: 0.7%), had a LoS shorter than 30 days. In-hospital mortality rate until 30 days post-admission was 4.7% (SD: 0.7%). In-hospital mortality rate for patients with a LoS longer than 30 days was 11.6% (SD: 2.4%).

Overall mortality rate at 30 days (both in-hospital and out-hospital) was 7.2% (SD: 0.8%). Mortality rate from admission to 30 days post-discharge was 8.4% (SD: 0.9%). Early post-discharge mortality rate was 2.7% (SD: 0.4%). For a full overview of mortality rates and discharge statistics, see table 4.2.

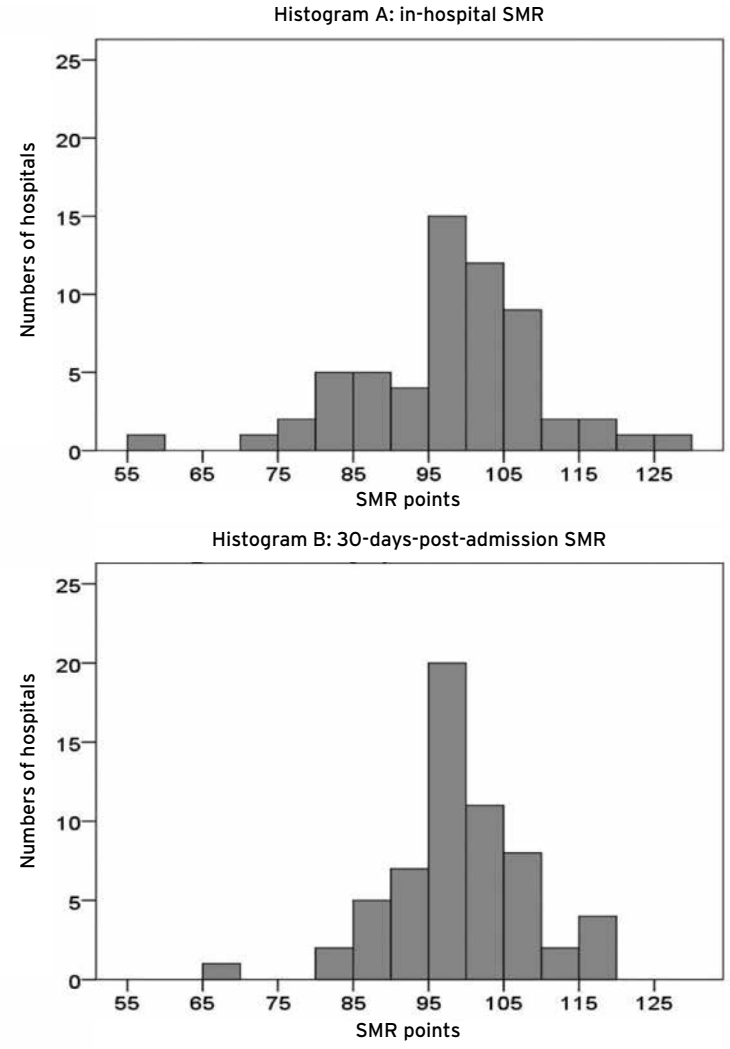
Table 4.2 Overview of crude mortality rates, transfer rates, and average length of hospital stay

Measure	Mean (SD)
In-hospital mortality rate (%)	4.9 (0.7)
Hospital mortality rate until 30 days after admission (%)	7.2 (0.8)
Hospital mortality rate until 30 days after discharge (%)	8.4 (0.9)
Length of hospital stay (days)	7.2 (0.7)
Admissions with length of hospital stay <30 days (%)	97.8 (0.7)
In-hospital mortality rate at 30 days after admission (%)	4.7 (0.7)
Early post-discharge mortality rate ($\dagger_{discharge}$-$\dagger_{admission+30days}$) (%)	2.7 (0.4)
In-hospital mortality rate for admissions >30 days (%)	11.6 (2.4)
Transfer rate (%)	9.4 (3.8)

Comparison of hospital SMRs based on different mortality timeframes

Figure 4.1 shows histograms of in-hospital SMRs, 30 days post-admission SMRs, and 30 days post-discharge SMRs. Between-hospital variability was less with 30 days post-admission SMRs and 30 days post-discharge SMRs. Figure 4.2 shows scatterplots indicating how in-hospital SMRs change when 30 days post-admission SMRs and 30 days post-discharge SMRs are used.

Figure 4.1 Distribution of hospitals according to in-hospital standardised mortality ratios (SMR), 30 days post-admission SMR, and 30 days post-discharge SMR



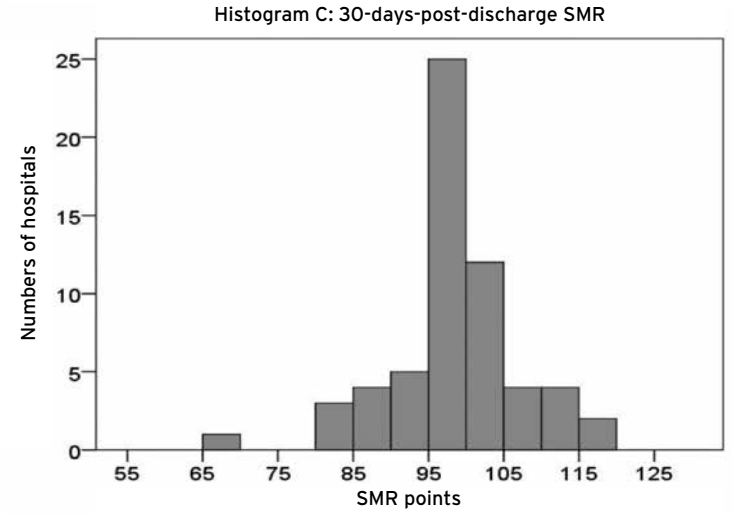


Table 4.3 and 4.4 show whether these different SMRs also lead to different judgment of individual hospitals. Based on in-hospital SMR, 17 hospitals performed better than expected (95% C.I. < 100) and 9 hospitals performed worse than expected (95% C.I. > 100). Using 30 days post-admission SMR, 20 hospitals were judged differently compared with in-hospital HSMR (Table 4.3).
With 30 days post-discharge SMR, 13 hospitals were categorized differently (Table 4.4).

Table 4.3 Classification according to 30 days post-admission standardised mortality ratio (SMR) compared with in-hospital SMR

Classification according to in-hospital SMR	Classification according to 30 days post-admission SMR		
	Worse than expected	Conforms to expected	Better than expected
Worse than expected	6	3	0
Conforms to expected	5	23	6
Better than expected	0	6	11

On the basis of the SMR and its 95% confidence interval, a hospital can be classified in three categories: better than expected, conforms to expected, and worse than expected. If the SMR is significantly above 100 or significantly below 100, the hospital is considered to have performed respectively worse or better than expected. If the SMR does not significantly differ from 100, the hospital's performance is considered to have conformed to expected. Twenty out of 60 hospitals were classified differently with the 30 days post-admission timeframe in comparison with in-hospital mortality.

Table 4.4 Classification according to 30 days post-discharge standardised mortality ratio (SMR) compared with in-hospital SMR

Classification according to in-hospital SMR	Classification according to 30 days post-discharge SMR		
	Worse than expected	Conforms to expected	Better than expected
Worse than expected	6	3	0
Conforms to expected	3	28	3
Better than expected	0	4	13

On the basis of the SMR and its 95% confidence interval, a hospital can be classified in three categories: better than expected, conforms to expected, and worse than expected. If the SMR is significantly above 100 or significantly below 100, the hospital is considered to have performed respectively worse or better than expected. If the SMR does not significantly differ from 100, the hospital's performance is considered to have conformed to expected. Thirteen out of 60 hospitals were classified differently with the 30 days post-discharge timeframe in comparison with in-hospital mortality.

Effect of discharge patterns on in-hospital SMR

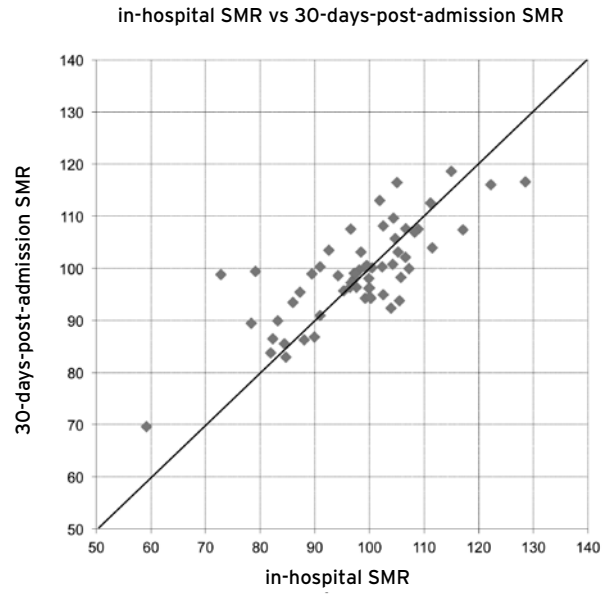
Table 4.5 shows the association between in-hospital SMR and early post-discharge mortality rates, LoS, and transfer rates. The in-hospital SMR had a positive correlation with LoS (Pearson correlation coefficient = 0.33 (P = 0.01)) and a negative correlation with early post-discharge mortality rates (Pearson correlation coefficient = -0.37 (P = 0.004)). The correlation between LoS and early post-discharge mortality rates was negative (Pearson correlation coefficient = -0.30 (P = 0.02)).

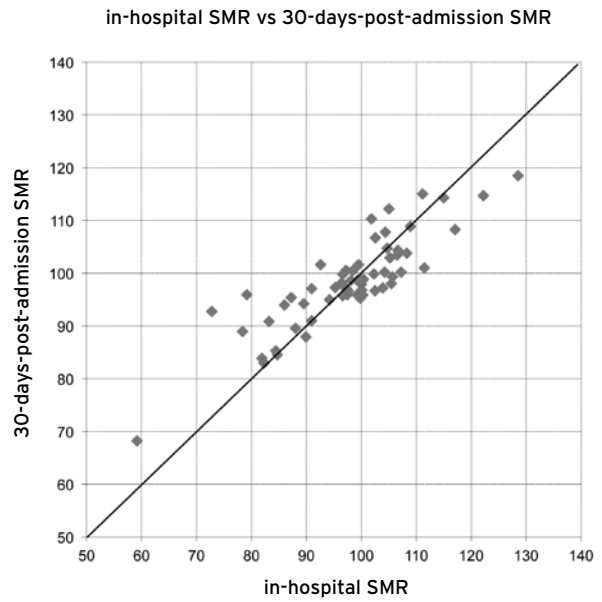
According to the used dataset 9.4% discharges (SD: 3.8%) were transferred to another medical institution. We noted that four hospitals had no recorded transfers. The correlation between in-hospital SMR and transfer rate was not statistically significant (Pearson correlation coefficient = -0.06 (P = 0.661)).

Table 4.5 Relations between in-hospital standardised mortality ratio (SMR) and early post-discharge mortality rate, transfer rate, and length of stay, and between length of stay and early post-discharge mortality

Relation	Pearson correlation coefficient	P value
In-hospital SMR and early post-discharge mortality	-0.37	0.004
In-hospital SMR and transfer rate	-0.06	0.66
In-hospital SMR and length of stay	0.33	0.01
Length of stay and early post-discharge mortality	-0.30	0.02

Figure 4.2 Scatterplots showing that for some individual hospitals standardised mortality ratio (SMR) changes if 30 days post-admission or 30 days post-discharge SMRs are used. The diagonal indicates the points where in-hospital SMR equals 30 days post-admission SMR (or 30 days post-discharge SMR)





DISCUSSION

This study examined the effect of applying mortality timeframes that include the post-discharge period on SMRs of individual hospitals. Compared with SMRs based on in-hospital mortality, we found that these timeframes resulted in differences in SMRs, and even altered judgments regarding the performance of individual hospitals. Furthermore, we found associations between in-hospital SMR, LoS, and early post-discharge mortality. Combining these findings suggests that SMRs based on in-hospital mortality may be subject to so-called 'discharge bias'.[23]

The presence of discharge bias is suggested by several observations in our analysis. First, we found an inverse relationship between in-hospital SMR and early post-discharge mortality implicating that lower in-hospital mortality may actually reflect higher post discharge mortality instead of the assumed higher degree of quality of care.

Second, a shorter average LoS is associated with lower in-hospital SMR. To be considered as 'better performing', hospitals with low LoS and with low in-hospital mortality should also have low or average post-discharge mortality. However, we found that the correlation between average LoS and early post-discharge mortality is negative, implicating that shorter average LoS is associated with higher post-discharge mortality. If a hospital decides to reduce LoS without changing the care delivered, more patients will die after discharge instead of during admission; as a consequence in-hospital SMR will decrease. This phenomenon may be increasingly important as the average LoS is consistently declining over the last decades[13, 14], and will most likely continue to decline in the future due to economic pressures on hospital-beds. Finally, the histograms in figure 4.1 show that the between-hospital SMR variability decreased when 30 days post-admission and 30 days post-discharge timeframes are used, suggesting that at least part of the variation of in-hospital SMR can be explained by mortality occurring shortly after discharge. Altogether, the influence of discharge bias may be substantial for individual hospitals, as a large number of hospitals (20 out of 60) were categorised differently when the 30 days post-admission timeframe is used.

Comparison with other studies

Our results are in accordance with previous work and underline the risk of discharge bias when using in-hospital mortality statistics. For example, Vasilevskis and colleagues (2009) studied ICU admissions and concluded

that variations in transfer rates and discharge timing appear to bias in-hospital SMR calculations[23]. In the present study, we found a statistically significant association between early post-discharge mortality and in-hospital SMRs but no statistically significant association between transfer rates and in-hospital SMRs. However, the fact that four hospitals did not record any transfers to other medical institutions at all, suggests that the quality of registration of this variable in our database is questionable, at least for some hospitals. Therefore, the lack of a statistically significant effect could also be due to a poor quality of registration of this variable. A recent study of Drye and colleagues (2012) concluded that for patients admitted with the diagnoses acute myocardial infarction, heart failure, and pneumonia, in-hospital mortality rates favour hospitals with shorter LoS and higher transfer rates compared with 30 days post-admission mortality rates[24]. They reported that a higher LoS was associated with higher in-hospital mortality. Our results are in line with these observations. Rosenthal et al. (2000) examined the relationship between in-hospital mortality and hospital discharge practices using data of 13,834 patients with congestive heart failure in the United States[25]. They found that the classification of hospitals as statistically significant outliers on the basis of their in-hospital SMRs was noticeably different from the classification based on 30 days post-admission SMRs. This observation may well suggest the presence of discharge bias. In addition to the previously mentioned studies, the present study is, to our knowledge, the first to include a broad hospital population taking into account 50 diagnoses with a higher a priori mortality risk. Therefore our results may be more applicable when studying hospital-wide performance, for example when using HSMRs to assess quality of care.

Limitations of study

The pseudonomised, administrative data used for this study have some limitations that need to be considered. First, approximately 10% of the pseudonomised admissions could not be linked to the population register and had to be excluded. Unfortunately, we had no other means to retrieve the post-discharge mortality of the 10% non-linkable admissions and therefore we do not know whether these discharges might have influenced the magnitude and direction of the difference between in-hospital SMR and post-discharge mortality. However, Statistics Netherlands considers this number of linkable admissions sufficient to perform statistical analysis[19]. Second, we could not determine whether the 50 analysed diagnostic groups, accounting for 80% of in-hospital mortality, also accounted for a high percentage of post-

discharge mortality. This is because admissions not belonging to the 50 analysed diagnostic groups were not linked to the population register, and therefore, post-discharge mortality could not be computed for these admissions. Third, because participation of hospitals in the HDR-database was on a voluntary basis, not all hospitals participated, potentially reducing the variation in hospital performance (especially, when poorly performing hospitals selectively decline to participate). However, it is likely that the included hospitals give a fair representation of all Dutch general hospitals, because the crude mortality rates of the excluded hospitals were similar to those of included hospitals. Finally, we had no means to determine whether a patient had been admitted to, or was transferred from a non-participating hospital. Consequently, the death of a transferred patient from or to a hospital that did not contribute to the HDR-database was only assigned to the admitting or referring hospital that participated in the HDR.

Implications of findings

Increasingly, pay-for-performance programmes and selective purchasing are based on outcomes rather than adherence to process variables. If hospital mortality – either at the aggregate level (HSMR) or by specialty, diagnosis or procedure – is used as a performance measure, it is essential to guard against bias and reduce the potential for 'gaming'. We found that between-hospital variability of in-hospital SMRs could be partly explained by differences in post-discharge mortality and LoS. This skews interpretation of quality of care against hospitals with longer LoS and lower post discharge mortality rates. Therefore, we recommend including post-discharge deaths in the mortality analyses by using timeframes that incorporate the early post-discharge period. Of course, mortality after discharge may also be affected by factors beyond hospital control, such as quality of out-patient care or quality of other referring and admitting hospitals. However, from a societal perspective this could be beneficial, as hospitals now have a stake in organizing adequate handover and post-discharge care. In addition, collection of post-discharge data is currently not routine and acquiring these data may be costly. Nevertheless, our study suggests that 30 days post-admission or 30 days post-discharge SMRs are less vulnerable to discharge bias than in-hospital SMRs and may therefore be preferable if SMRs are to be used for assessment of hospital performance.

30 days post-admission mortality vs. 30 days post-discharge mortality

Whether to use a 30 days post-admission timeframe or a 30 days post-discharge timeframe is still under debate. The major advantage of a 30 days post-admission timeframe is the fixed window of time in which care is measured. The timeframe is equal for all hospitals whatever their discharge policy or their opportunities to reduce the length of stay such as the near vicinity of palliative care centres or other more specialised hospitals. However, patients dying in the hospital after 30 days of admission will be mistakenly regarded as survivors, introducing a potential gaming element for hospitals (an extreme example would be the incentive to keep patients with a poor prognosis alive until at least 30 days after admission). With the 30 days post-discharge timeframe these patients are also analysed, however the timeframe of measurement is no longer fixed. A more elegant method would be to determine the best timeframe per diagnosis or procedure. For example, for surgical procedures it is common to use a 30 days post-admission timeframe, but for some diagnoses such as pneumonia, a longer timeframe may be preferable. Also, sometimes a combination of timeframes is used; for example the 30 days post-admission timeframe for patients discharged within 30 days combined with the in-hospital timeframe for patients admitted longer than 30 days. This has been successfully applied in the EuroSCORE, a tool to predict operative mortality for patients undergoing cardiac surgery[26]. Further research is required to determine the optimal window of time for every specific diagnosis. Based on our findings and the literature the 30 days post-admission timeframe in combination with the in-hospital timeframe for patients admitted longer than 30 days may best balance the risk of discharge bias and maintain the advantage of a fixed timeframe.

Conclusions

In conclusion, selecting mortality timeframes that include the post-discharge period changes SMRs of individual hospitals and affects performance judgments. Furthermore, short LoS was associated with low in-hospital mortality but higher post-discharge mortality. These findings suggest that incorporating early post-discharge mortality in the SMR will reduce the effect of discharge bias.

REFERENCE LIST

1 Jarman B, Galt S, Alves B, Hider A, Dolan S, Cook A, et al. Explaining differences in English hospital death rates using routinely collected data. *BMJ* 1999; 318:1515-20.

2 Jarman B, Pieter D, van der Veen AA, Kool RB, Aylin P, Bottle A, et al. The hospital standardised mortality ratio: a powerful tool for Dutch hospitals to assess their quality of care? *Qual Saf Health Care* 2010; 19:9-13.

3 Ben-Tovim D, Woodman R, Harrison JE, Pointer S, Hakendorf P, Henley G. Measuring and reporting mortality in hospital patients. Canberra: Australian Institute of Health and Welfare, 2009. (AIHW Cat. No. HSE 69.)

4 Canadian Institute for Health Information. HSMR Technical notes, updated September 2012. Ottawa 2012, http://www.cihi.ca/CIHI-ext-portal/pdf/internet/ HSMR_TECH_NOTES_201202_en.pdf

5 Pitches DW, Mohammed AM, Lilford RJ. What is the empirical evidence that hospitals with higher-risk adjusted mortality rates provide poorer quality care? A systematic review of the literature. *BMC Health Serv Res* 2007; 7:91.

6 Shahian DM, Wolf RE, Iezzoni LI, Kirle, L, Normand SLT. Variability in the Measurement of Hospital-wide Mortality Rates. *N Engl J Med* 2010; 363:2530-9.

7 Tu Y-K, Gilthorpe MS. The most dangerous hospital or the most dangerous equation? *BMC Health Serv Res* 2007; 7:185-9.

8 Powell H, Li LL, Heller RF. Accuracy of administrative data to assess comorbidity in patients with heart disease: an Australian perspective. *J Clin Epidemiol* 2001; 54:687-93.

9 Van Gestel YRBM, Lemmens EPP, Lingsma HF, de Hingh IHJT, Rutten HJT, Coebergh JW. The Hospital Standardised Mortality Ratio Fallacy, a narrative review. *Med Care* 2012; 50:662-7

10 Van den Bosch WF, Kelder JC, Wagner C. Predicting hospital mortality among frequently readmitted patients: HSMR biased by readmission. *BMC Health Serv Res* 2011; 11:57.

11 Mohammed MA, Deeks JJ, Girling A, Rudge G, Carmalt M, Stevens JA, et al. Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals. *BMJ* 2009; 338:b780.

12 Pouw ME, Peelen LM, Lingsma HF, Pieter D, Steyerberg E, Kalkman CJ, Moons KGM. Hospital standardized mortality ratio: consequences of adjusting hospital mortality with indirect standardization. *PLOS One* 2013; 8(4): e59160

13 Borghans I, Heijink R, Kool T, Lagoe RJ, Westert GP: Benchmarking and reducing length of stay in Dutch hospitals. *BMC Health Serv Res* 2008, 8:220.

14 Slobbe LCJ, Onyebuchi AA, de Bruin A, Westert GP. Mortality in Dutch hospitals: Trends in time, place and cause of death after admission for myocardial infarction and stroke. An observational study. *BMC Health Serv Res* 2008; 4; 8:52.

15 Yu P, Chang DC, Osen HB, Talamani MA. NSQIP reveals significant incidence of death following discharge. *J Surg Res* 2011, 170(2): e217-24.

16 Metersky ML, Waterer G, NSA W, Bratzler DW. Predictors of in-hospital vs postdischarge mortality in pneumonia. *Chest* 2012, 142(2): 476-81.

17 Campbell MJ, Jacques RM, Fotheringhma J, Maheswaran R, Nicholl J. Developing a summary hospital mortality index: retrospective analysis in English hospitals over five years. *BMJ* 2012, 344:e1001

18 Bratzler DW, Normand SL, Wang Y, O'Donnell WJ, Metersky M, Han LF, Rapp MT, Krumholz HM. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. *PLOS One* 2011, 6(4): e17401.

19 de Bruin A, Kardaun J, Gast F, de Bruin E, van Sijl M, Verweij G. Record linkage of hospital discharge register with population register: Experiences at Statistics Netherlands. *Statistical Journal of the United Nations* ECE 21 (2004) 23–32

20 Israels A, van der Laan J, de Bruin A, Ploemacher J, Verweij G.HSMR 2010, methodological report. Statistical Methods (201122). http://www.cbs.nl/en-GB/menu/themas/gezondheid-welzijn/publicaties/artikelen/archief/2011/2011-hsmr2010-report.htm

21 Elixhauser A, Andrews RM, Fox S. Clinical classifications for health policy research: Discharge statistics by principal diagnosis and procedure. Provider Studies Research Note 17. Rockville, MD: Agency for Health Care Policy and Research 1993, AHCPR Publication No. 93-0043

22 Lawless, JF and Singhal K. Efficient screening of nonnormal regression models. *Biometrics* 1978; 34, 318-327.

23 Vasilevskis EE, Kuzniewicz MW, Dean ML, Clay T, Vittinghoff E, Rennie DJ, et al. Relationship between discharge practices and intensive care unit in-hospital mortality performance: evidence of a discharge bias. *Med Care* 2009; 47(7): 803-12.

24 Drye EE, Normand SL, Wang Y, Ross JS, Schreiner GC, Han L, et al. Comparison of hospital risk-standardized mortality rates calculated by using in-hospital and 30-day models: an observational study with implications for hospital profiling. *Ann Intern Med* 2012; 156:19-26

25 Rosenthal GE, Baker DW, Norris DG, Way LE, Harper DL, Snow RJ. Relationship Between In-Hospital and 30-Day Standardized Hospital Mortality: Implications for Profiling Hospitals. *Health Serv Res* 2000, 34:7

26 Roques F, Nashef SAM, Michel P, Gauducheau E. de Vincentiis C, Baudet E, Cortina J, et al. Risk factors and outcome in European cardiac surgery: analysis of the Euro-SCORE multinational database of 19030 patients. *Eur J Cardio-thorac Surg* 1999;15:816-23.

CHAPTER 5

Standardised mortality ratios of specialised hospitals: the influence of referral bias

Pouw ME, Peelen LM, Rinkel GJE,
Moons KGM, Kalkman CJ

ABSTRACT

Objectives

To assess to what extent standardised mortality ratios (SMRs) for 'aneurysmal subarachnoid haemorrhage' (ASAH) and 'traumatic brain injury' (TBI) in tertiary care neurosurgical centers differ from regional hospitals and examined the influence of referrals on SMRs.

Design

Retrospective analysis of routinely collected hospital data comparing observed deaths with deaths predicted with a casemix adjustment method for the clinical domains 'traumatic brain injury' and 'aneurysmal subarachnoid haemorrhage'.

Setting

Six tertiary care neurosurgical centers and 54 regional hospitals in the Netherlands.

Participants

2,536 admissions with ASAH, and 34,468 admissions with TBI.

Main outcome measures

In-hospital standardised mortality ratio and 30 days post-admission standardised mortality ratio.

Results

According to the SMRs, all hospitals performing worse than expected in the treatment of both ASAH and TBI were tertiary care centers. None of the tertiary care centers performed better than expected. The expected mortalities of ASAH and TBI referrals to tertiary care centers and regional hospitals were practically similar.

Conclusions

Tertiary care centers most likely have falsely high SMRs for ASAH and TBI due to lack of important clinical predictors. Furthermore, even with SMRs including the early post discharge mortality, referral bias is possible. To further reduce referral bias, the calculated expected mortality should also take into account the phase of illness to distinguish 'step-up' referrals from 'step-down' referrals.

INTRODUCTION

Standardised mortality ratios (SMRs) are a commonly used instrument to assess and compare hospital performance.[1-4] Given the implications of SMRs for hospitals in an era of pay-for-performance and public reporting, various studies have been conducted to improve the validity of SMRs.[5-7] Patients with aneurysmal subarachnoid haemorrhage (ASAH) and severe traumatic brain injury (TBI) are typically treated in tertiary care neurosurgical centers. Although outcomes of patients with ASAH and TBI have improved, mortality and disability rates remain high,[8] drawing growing attention to hospital performance in these clinical domains.[9-12]

Analyses based only on in-hospital deaths (in-hospital SMRs) are vulnerable to discharge bias and referral bias.[13,14] *Discharge bias* can occur if a patient dies outside the hospital shortly after discharge, in which case mortality is not accounted for in the calculation of the in-hospital SMR of the discharging hospital. If different hospitals have different discharge policies, this may hinder hospital comparisons. To reduce the effect of discharge bias, several studies have suggested that incorporation of early post-discharge mortality in the calculation of SMRs may reduce this type of bias.[13,14] In addition, *referral* bias can occur: if a patient with ASAH is referred to a tertiary hospital and subsequently dies, the in-hospital SMR of the receiving hospital is affected negatively, whereas the in-hospital SMR of the referring hospital is affected positively (as the patient was discharged alive). Referral bias thus influences 1) so-called 'step-up' referrals, where patients are referred to a more specialised hospital to receive specialised care or interventions, but also 2) 'step-down' referrals, where patients are referred to a less specialised hospital because they may no longer need specialised care. Differences in expected mortality between these two types of referrals create referral bias if there are differences in distribution of 'step-up' and 'step-down' referrals between hospitals.

In the Netherlands, only few hospitals, all tertiary care neurosurgical centers, are capable and appointed to treat patients with ASAH or major TBI. Patients with these diagnoses, initially admitted in other hospitals, are typically referred to these centers as soon as possible. After the intervention, patients may be transferred back to non-specialised hospitals for further recovery and treatment. It is unclear to which extent such different referral patterns in these two major neurologic diagnoses influence performance measures such as SMRs.

We assessed to what extent SMRs for ASAH and TBI in tertiary care neurosurgical centers differ from regional hospitals and examined the influence of referrals on SMRs.

METHODS

Data

Data were obtained from the National Hospital Discharge Register (HDR). Dutch Hospital Data (DHD), holder of the HDR, gave permission to use their database to perform this study. The HDR contains discharge data of Dutch hospitals and comprises patient characteristics such as age and gender as well as standard medical variables such as date of admission, date of discharge, diagnoses, and comorbidities. The HDR follows the ICD9-CM (international classification of diseases, 9th version, clinical modification) to register discharge diagnoses. To acquire information on deaths that occurred after discharge from hospital, Statistics Netherlands (www.CBS.nl) linked records from the Hospital Discharge Register to the Dutch population register.[15]
Of the 84 hospitals participating in the DHD, we analysed the 3-year SMR (2008-2010) for 60 hospitals. Six hospitals were specialised, tertiary care neurosurgical centers and 54 were non-specialised regional hospitals. The other 24 hospitals were not included in our analysis for various reasons, such as insufficient participation in the HDR, no inpatient admissions, insufficient data quality, and no permission to publish. A detailed explanation of these reasons is described in the methodological report of the Dutch HSMR 2010.[16]
We used the 3-year SMR because this is common practice in The Netherlands.

The HDR registers whether an admission is a 'referral' from another hospital, but does not contain explicit information on whether a referral was a 'step-up' or 'step-down' referral. Therefore, we chose two clinical diagnoses for which the referral type could straightforwardly be identified given the organization of care for these diagnoses in the Netherlands, namely aneurysmal subarachnoid haemorrhage (ASAH) and traumatic brain injury (TBI).[17]
We assumed that the majority of referrals to tertiary care centers are 'step-up' referrals and the majority of referrals to regional hospitals 'step-down' referrals from tertiary care centers.

Risk adjusted mortality models

For both primary diagnoses ASAH and TBI (coded using Clinical Classification System, CCS [18]), a case mix adjustment model for expected mortality was developed according to the methodology of Statistics Netherlands,[16] which is in line with commonly used methods for SMR calculation.[1, 2] The model is a logistic regression model with mortality as the dependent variable and age, sex, socio-economic status, sub-classification of main diagnosis, urgency of admission, comorbidities according to the Charlson Comorbidity Index,[19] source of admission, and month of admission as predictor variables. The regression model was used to calculate the expected probability of mortality for each patient with this diagnosis.

For both ASAH as well as TBI, we estimated two case mix adjustment models, the first using in-hospital mortality as the outcome, and the second using 30-days-post admission mortality as the dependent variable. All models were estimated in R version 2.15 and are available on request.

Standardised mortality ratios

The SMR is the ratio of the observed number of deaths and the expected number of deaths as calculated with the case mix adjustment model and then multiplied by the value of 100. If the 95% confidence interval of the SMR included the reference value of 100, hospital mortality was considered 'as expected'. If the lower level of the 95% confidence interval was above 100, the mortality was higher than expected, and if the upper limit of the 95% confidence interval was lower than 100, the mortality was lower than expected.

Comparing SMRs of tertiary care centers with regional hospitals

For both diagnoses ASAH and TBI, we calculated the in-hospital SMR and the 30-days-post-admission SMR for each hospital. Hospitals were classified into three groups (conform expected, better than expected, worse than expected) based on the 95% confidence interval of their SMRs.

Furthermore, we grouped the hospitals into tertiary care centers and regional hospitals and calculated and compared the SMR of the two hospital groups.

Comparing mortality of referrals between tertiary care centers and regional hospitals

Also, we zoomed in onto patient characteristics of admissions that were registered as 'referred from another hospital' (referrals), using means (standard deviation) or median (interquartile range) for normally and non-normally distributed variables respectively, and numbers and percentages for categorical variables. Mortality figures of referrals of tertiary care centers were compared with those of regional hospitals. Additionally, the effect of mortality of referrals on the total SMR number was analysed for both groups of hospitals.

RESULTS

Patients and hospital characteristics

Table 5.1 shows patient characteristics for the tertiary care centers and regional hospitals. In 2008-2010 the 60 included hospitals had 2,536 admissions with ASAH, and 34,468 admissions with TBI. The average numbers of admissions in tertiary care centers were respectively ten times and twice higher than in regional hospitals.

Table 5.1 Baseline characteristics of admissions diagnosed with aneurysmal subarachnoid haemorrhage and traumatic brain injury

Baseline characteristics	Aneurysmal subarachnoid haemorrhage		Traumatic brain injury	
	Regional hospitals	Tertiary care centers	Regional hospitals	Tertiary care centers
No. of hospitals	54	6	54	6
Admissions Referrals	1,111 199 (17.9%)	1,425 322 (22.6%)	28,234 564 (2.0%)	6,234 249 (4.0%)
Average Age (in years)	61.9	56.2	42.4	41.7
Male	444 (40.0%)	471 (33.1%)	16,189 (57.3%)	3,840 (61.6%)
Urgent admission	886 (79.7%)	1,222 (85.8%)	26,886 (95.2%)	5,982 (96.0%)
Average No. of comorbidities per admission	0.16	0.14	0.07	0.07
Average No. of admissions	20.6	237.5	522.9	1039.0
Average length of stay	6.8 days	15.7 days	3.6 days	6.3 days

Comparing SMRs of tertiary care centers with regional hospitals

Table 5.2A and 5.2B show the in-hospital SMR and 30-days-post-admission SMR when hospitals are grouped into tertiary care centers and regional hospitals. The grouped in-hospital SMR of tertiary care centers is significantly higher than 100 for both ASAH and TBI. The grouped in-hospital SMR of regional hospitals is lower than 100 for ASAH and significantly lower than 100 for TBI. The grouped 30-days-post-admission SMR for tertiary care centers is still significantly higher than 100 for TBI but not for ASAH.

Focusing on the classification of individual hospitals (table 5.3), all hospitals performing worse than expected for the treatment of both SAH and TBI were tertiary care centers. None of the tertiary care centers were performing better than expected.

Table 5.2A Observed mortalities, expected mortalities, and SMRs of SAH admissions, stratified in regional hospitals and tertiary care centers and subsequently in referred and non-referred admissions

Subarachnoid haemorrhage	Regional hospitals (n = 54)			Tertiary care centers (n = 6)		
	Non-referred admissions	Referred admissions	All admissions	Non-referred admissions	Referred admissions	All admissions
Admissions	912 (82%)	199 (18%)	1111	1103 (77%)	322 (23%)	1425
Average length of stay (days)	5.1	14.5	6.8	16.2	13.9	15.7
In-hospital mortality						
In-hospital mortality rate[1]	24.8%	7.5%	21.7%	23.2%	13.0%	20.9%
In-hospital expected mortality rate	26.3%	11.5%	23.7%	20.3%	11.5%	18.3%
30-days-post admission mortality						
30-days mortality rate[2]	33.3%	8.5%	28.9%	24.2%	14.9%	22.1%
30-days expected mortality rate	30.8%	12.3%	27.5%	23.8%	12.8%	21.4%
SMR						
in-hospital SMR	94 (82 - 107)	66 (34 - 109)	92 (91 - 104)	115 (101 - 130)	113 (82 -153)	114 (102 - 128)
30-days SMR	108 (96 - 121)	70 (41 - 111)	105 (94 -117)	102 (90 - 115)	116 (86 - 154)	104 (92 - 116)

1 In-hospital mortality: mortality during entire hospital admission.
2 30 days post admission mortality: mortality from admission to 30 days after admission.

Table 5.2B Observed mortalities, expected mortalities, and SMRs of TBI admissions, stratified in regional hospitals and tertiary care centers and subsequently in referred and non-referred admissions

Traumatic brain injury	Regional hospitals (n = 54)			Tertiary care centers (n = 6)		
	Non-referred admissions	Referred admissions	All admissions	Non-referred admissions	Referred admissions	All admissions
Admissions	27670 (98%)	564 (2%)	28234	5985 (96%)	249 (4%)	6234
Average length of stay (days)	3.4	13.2	3.6	6.1	10.9	6.3
In-hospital mortality						
In-hospital mortality	1.6%	6.2%	1.6%	5.7%	9.2%	5.8%
In-hospital mortality rate[1]	2.1%	7.5%	2.2%	4.3%	6.5%	4.4%
30-days-post-admission mortality						
30-days mortality rate[2]	2.1%	6.2%	2.2%	5.9%	12.9%	6.2%
30-days expected mortality rate	2.7%	8.7%	2.8%	4.8%	7.9%	4.9%
SMR						
in-hospital SMR (95% confidence interval)	73 (66 - 79)	83 (58 - 115)	73 (67 - 80)	132 (119 - 147)	142 (90 - 213)	133 (120 - 147)
30-days SMR (95% confidence interval)	79 (73 - 86)	67 (46 - 95)	78 (72 - 85)	125 (112 - 138)	163 (111 - 230)	127 (115 - 140)

1 In-hospital mortality: mortality during admission.
2 30 days post admission mortality: mortality from admission to 30 days after admission.

Comparing mortality of referrals between tertiary care centers and regional hospitals

As depicted in table 5.1 the proportion of referred admissions was higher in tertiary care centers than in regional hospitals. The crude in-hospital mortality and the crude 30-days-post-admission mortality of referred admissions were both higher in tertiary care centers than in regional hospitals for ASAH (tables 5.2A and 5.2B).

Table 5.3 Classification of tertiary care centers and regional hospitals according to in-hospital SMR and 30-days-post-admission SMR

Classification of hospitals according to in-hospital SMR

	Subarachnoid haemorrhage		Traumatic brain injury	
	Regional hospitals	Tertiary care centers	Regional hospitals	Tertiary care centers
Worse than expected	0	1	0	4
Conform expected	53	5	47	2
Better than expected	1	0	7	0

Classification of hospitals according to 30 days post-admission SMR

	Subarachnoid haemorrhage		Traumatic brain injury	
	Regional hospitals	Tertiary care centers	Regional hospitals	Tertiary care centers
Worse than expected	0	0	0	4
Conform expected	54	6	49	2
Better than expected	0	0	5	0

'Worse than expected': 95% confidence interval of SMR greater than 100
'Conform expected': 95% confidence interval of SMR includes 100
'Better than expected': 95% confidence interval of SMR lower than 100

For ASAH the expected in-hospital mortality of referred admissions was 11.5% both in tertiary care centers and regional hospitals. The expected 30-days-post-admission mortality of referred ASAH admissions in tertiary care centers was 12.8% and in regional hospitals 12.3%.

For TBI the expected in-hospital mortality of referred admissions in tertiary care centers was 6.5% vs. 7.5% in regional hospitals. Expected 30-days-post-admission mortality was 7.9% versus 8.7% respectively.

For ASAH, the in-hospital SMR and 30-days-post admission SMR of tertiary care centers are almost similar to the corresponding SMRs without the inclusion of referrals, respectively 114 vs. 115 and 104 vs. 102 (table 5.2A). For regional hospitals, we also saw no significant difference between the SMRs with and without referrals.

For TBI, SMRs including and excluding referrals did not differ for both tertiary care centers and regional hospitals (table 5.2B).

DISCUSSION

We found that SMRs for ASAH and TBI in tertiary care neurosurgical centers were considerably higher than in regional hospitals. In our study, all hospitals with significantly higher SMRs – classified as performing 'worse than expected' – were tertiary care centers. The higher SMRs in tertiary care centers may be explained by a higher-than-expected observed mortality (which would imply that tertiary care centers are indeed performing worse than regional hospitals), or an underestimation of expected mortality, which could be estimated erroneously due to the following reasons.

First, information may be lacking on *severity* of illness at admission. For ASAH and TBI indicators of severity of injury (parameters of level of consciousness) are important predictors for mortality[20-22]. This information is typically not collected in administrative databases, which serve as the basis for calculation of expected mortality. Insufficient casemix adjustments lead to over-, and underestimation of the expected mortality. In particular, specialised hospitals such as tertiary care centers, admitting severely ill patients, are likely put at a disadvantage. The higher SMRs we observed in tertiary care centers contrast with studies that reported better outcomes in specialised hospitals.[23, 24] For example, Dubose et al. concluded that patients with severe TBI treated in level-1 trauma centers had better survival rates and outcomes than those treated in level-2 trauma centers.[23] Garwe et al. found a survival benefit among patients transferred to specialised trauma centers compared to those remaining in non-specialised trauma centers.[24] However, these studies adjusted for patient-level risk factors such as Glasgow coma scales and Injury Severity Score.

Second, information regarding the *phase of illness* (acute phase versus stable phase after initial treatment) is lacking in administrative databases. ASAH and TBI referrals to tertiary care centers are typically in the acute phase of their disease with a high probability of complications and death (step-up), whereas referrals to regional hospitals tend to be in a more stable phase of their disease with in general less probability of complications (step-down). The lack of distinction between 'step-up' and 'step-down' referrals within prediction models averages out any potential differences between referrals to tertiary care centers and to regional hospitals. Indeed, we found that expected mortality rates for referred TBI admissions, as well as for referred ASAH admissions, are nearly similar for tertiary care centers and regional hospitals (tables 5.2A and 5.2B). So, although in clinical practice step-up and step-down referrals have different probabilities of dying – partly due to the phase of illness – this is not reflected in the calculated expected mortality rates.

Furthermore, our results suggest that using in-hospital mortality as a basis for SMR calculation increases the potential of referral bias compared to using 30-days-post- admission mortality. For patients with ASAH initially admitted to a regional hospital, 30-days-post-admission mortality rate is considerably higher than in-hospital mortality rate (33.3% vs 24.8%). This can be explained as follows: ASAH patients are referred (step-up referral) as soon as possible to a tertiary care center for definitive treatment (coiling or clipping). For the receiving tertiary care center, in-hospital mortality and 30-days-post- admission mortality are similar (13.0% vs 14.9%), indicating that if patients die after such specialised treatment, most of them do so within the receiving tertiary care center. In contrast, for the referring regional hospital, SMR will be low when calculated based on in-hospital mortality, but higher when mortality occurring in the tertiary center is included in the SMR based on 30-day mortality. Thus using 30-days-post- admission mortality addresses aspects of referral bias with respect to observed mortality (but not with respect to expected mortality).

Our study is, to our knowledge, the first that addresses referral bias by comparing mortality of referrals to tertiary care centers (step-up) with referrals to regional hospitals (step-down). Several studies examined the influence of referrals on hospital performance evaluation.[25-28] However, these studies had a different research objective than ours, comparing outcomes of referred patients with directly admitted patients. They concluded that, even with a sophisticated casemix adjustment method using precise severity of illness information data, patients transferred to an ICU of a tertiary care center had worse outcomes than those who were directly admitted. Our findings suggest that for the diagnoses SAH and TBI information about the phase of illness

is necessary to adjust for differences in 'step-up' and 'step-down' referrals reducing referral bias between hospitals.

There are limitations to this study. First, the administrative dataset did not contain explicit information whether a referral was 'step-up' or 'step-down'. However, in the Netherlands organisational structure of care for patients with TBI and ASAH is such that it allowed us to assume that most of the ASAH and TBI referrals to tertiary care centers are 'step-up' referrals and those to regional hospitals 'step-down' referrals. Whether the number of referrals is underregistered is a second issue. In the HDR database, on average 23% of all SAH admissions of tertiary care centers were referrals from other hospitals. This seems rather low compared with estimates from existing specialised ASAH databases. As a higher proportion of referrals will increase the probability of referral bias, the need for additional information concerning the phase of illness is more pressing. Third, we could not determine whether the inclusion of clinical predictors (e.g. GCS on admission, Injury Severity Score, mechanisms of injury, and Hunt and Hess Stroke scale) in the prediction models for ASAH and TBI would indeed alter the SMRs of tertiary care centers and regional hospitals, as these were not available. However, in tertiary care centers these clinical variables are the strongest predictors of mortality and long term outcomes after stroke.[29] Finally, because participation in the HDR was on a voluntary basis, not all hospitals participated, potentially altering the differences in SMR between tertiary care centers and regional hospitals.

Despite the limitations described, and the limited influence of referrals on the overall SMR in our study, our findings suggest that tertiary care centers are put to a disadvantage with the current SMR calculation due to lack of information about the severity of illness and the presence of referral bias between tertiary care centers and regional hospitals. We found almost equal expected mortalities for referrals to tertiary care centers and regional hospitals, which is not in line with the assumption that referrals in the acute phase of illness ('step-up' referrals) have a higher expected mortality than referrals in a more stable phase of illness ('step-down' referrals). Hence, admitting 'step-up' referrals can adversely affect a hospital's SMR, while accepting 'step-down' referrals can positively affect its SMR.

Referral bias and its consequences for SMR calculations are of course not restricted to hospitals specialised in acute neurological diseases, but may also occur in hospitals specialised in other acute diseases such as heart centers. The extent to which this leads to erroneous SMR estimations, depends on various aspects, such as the percentage of referrals and the importance of clinical indicators for the severity and phase of illness that are currently not in the SMR prediction models. Given these restrictions, it is questionable whether SMRs can be used

to benchmark specialised hospitals against non-specialised hospitals, because they cater to different levels of severity and different phases of the disease.

In conclusion, our findings suggest that tertiary care centers most likely have falsely high SMRs for SAH and TBI. The lack of (known) important clinical predictors in the computation of the expected mortality most likely puts tertiary care centers at a disadvantage as they will receive more severely ill admissions, either referred or directly admitted. Furthermore, the inclusion of the post-discharge period in the calculation of SMR addresses some aspects of referral bias – especially regarding the observed mortality – but it still neglects features concerning the expected mortality. To compare tertiary care centers with non-specialised regional hospitals important (clinical) predictors should be included in the prediction model. To further reduce referral bias, the calculated expected mortality should also take into account the phase of illness to distinguish 'step-up' referrals from 'step-down' referrals.

REFERENCE LIST

1 Jarman B, Galt S, Alves B, Hider A, Dolan S, Cook A, et al. Explaining differences in English hospital death rates using routinely collected data. BMJ 1999;318:1515-20

2 Jarman B, Pieter D, et al. The hospital standardised mortality ratio: a powerful tool for Dutch hospitals to assess their quality of care? Qual Saf Health Care 2010; 19:9-13

3 Ben-Tovim D, Woodman R, Harrison J, et al. Measuring and reporting mortality in hospital patients. Canberra: Australian Institute of Health and Welfare, 2009. (AIHW Cat. No. HSE 69.)

4 Alexandrescu R, Bottle A, Hua Jen M, Jarman B, Aylin P. The US hospital standardised mortality ratio: Retrospective database study of Massachusetts hospitals. *JRSM Open*. 2015;6(1):2054270414559083. doi:10.1177/2054270414559083.

5 van den Bosch WF, Kelder JC, Wagner C. Predicting hospital mortality among frequently readmitted patients: HSMR biased by readmission. *BMC Health Serv Res* 2011; 11:57.

6 Pouw ME, Peelen LM, Lingsma HF, Pieter D, Steyerberg E, Kalkman CJ, Moons KGM. Hospital standardized mortality ratio: consequences of adjusting hospital mortality with indirect standardization. *PLOS ONE* 2013; 8(4): e59160

7 Siregar S, Pouw ME, Moons KGM, et al. The Dutch Hospital Standardised Mortality Ratio (HSMR) method and cardiac surgery: benchmarking in a national cohort using hospital administration data versus a clinical database. *Heart* 2014; 100(9):702-710. doi:10.1136/heartjnl-2013-304645.

8 Nieuwkamp DJ, Setz LE, Algra A, et al. Changes in case fatality of aneurysmal subarachnoid haemorrhage over time, according to age, sex, and region: a meta-analysis. *Lancet Neurol* 2009; 8:635–642.

9 Pandey AS, Gemmete JJ, Wilson TJ, et al. High subarachnoid hemorrhage patient volume associated with lower mortality and better outcomes. *Neurosurgery* 2015; 77:462–470.

10 Prabhakaran S, Fonarow GC, Smith EE, et al. Hospital case volume is associated with mortality in patients hospitalized with subarachnoid hemorrhage. *Neurosurgery* 2014; 75:500–508.

11 Nathens AB, Cryer HG, Fildes J. The American College of Surgeons Trauma Quality Improvement Program. *The Surgical clinics of North America* 2012 Apr; 92(2):441–454.

12 Sharma S1, de Mestral C, Hsiao M, Gomez D, Haas B, Rutka J, Nathens AB. Benchmarking trauma center performance in traumatic brain injury: the limitations of mortality outcomes. *J Trauma Acute Care Surg* 2013 Mar;74(3):890-4.

13 Pouw ME, Peelen LM, Moons KGM, Kalkman CJ, Lingsma HF. Including post-discharge mortality in calculation of hospital standardised mortality ratios: retrospective analysis of hospital episode statistics. *BMJ* 2013; 347: f5913

14 Lamarche-Vadel A, Ngantcha M, Le Pogam MA, Ghosn W, Grenier C, Meyer L, Rey G. Hospital comparisons based on mortality: revisiting the choice of postadmission timeframe and evaluating the contribution of cause-of-death data, France, 2009. *Med Care* 2015 Aug;53(8):736-42.

15 De Bruin A, de Bruin EI, Gast A, Kardaun JWPF, van Sijl M, Verweij GCG, Koppeling van LMR- en GBA-gegevens. Methode, resultaten en kwaliteitsonderzoek. CBS Voorburg/Heerlen, 2003. http://www.cbs.nl/NR/rdonlyres/BAF2D6C5-7A77-4771-859B-338CC2F6F589/0/koppelingLMGBAgegevens1203.pdf

16 Israels A, van der Laan J, de Bruin A, Ploemacher J, Verweij G.HSMR 2010, methodological report. Statistical Methods (201122). http://www.cbs.nl/en-GB/menu/themas/gezondheid-welzijn/publicaties/artikelen/archief/2011/2011-hsmr2010-report.htm

17 Rinkel GJE, van Dijk JMC, Slooter AJC, et al. Richtlijn Subarachnoïdale bloeding. Diagnostiek en behandeling van een subarachnoïdale bloeding ten gevolge van een gebarsten cerebraal aneurysma 2013, https://nvic.nl/sites/default/files/Richtlijn%20SAB%20-%20definitief%202013.pdf

18 Elixhauser A, Andrews RM, Fox S. Clinical classifications for health policy research: Discharge statistics by principal diagnosis and procedure. Provider Studies Research Note 17. Rockville, MD: Agency for Health Care Policy and Research 1993, AHCPR Publication No. 93-0043 http://www.hcup-us.ahrq.gov/toolssoftware/icd_10/ccs_icd_10.jsp

19 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987; 40:373–83.

20 Chamoun RB, Robertson CS, Gopinath SP: Outcome in patients with blunt head trauma and a Glasgow Coma Scale score of 3 at presentation. *J Neurosurg* 2009; 111:683–687.

21 Grote S, Böcker W, Mutschler W, Bouillon B, Lefering R: Diagnostic value of the Glasgow Coma Scale for traumatic brain injury in 18,002 patients with severe multiple injuries. *J Neurotrauma* 2011; 28:527–534

22 Demetriades D, Kuncir E, Velmahos GC, Rhee P, Alo K, Chan LS. Outcome and Prognostic Factors in Head Injuries With an Admission Glasgow Coma Scale Score of 3. *Arch Surg* 2004; 139(10):1066-1068.

23 Dubose JJ, Browder T, Inaba K, et al. Effect of trauma center designation on outcome in patients with severe traumatic brain injury. *Arch Surg* 2008; 143(12):1213-1217

24 Garwe T, Cowan LD, Neas B, et al. Survival benefit of transfer to tertiary trauma centers for major trauma patients initially presenting to nontertiary trauma centers. *Acad Emerg Med* 2010; 17(11):1223-32. doi: 10.1111/j.1553-2712.2010.00918.x.

25 Combes A, Luyt CE, Trouillet JL, et al. Adverse effect on a referral intensive care unit's performance of accepting patients transferred from another intensive care unit. *Crit Care Med* 2005 Apr;33(4):705-10.

26 Rosenberg AL, Hofer TP, Strachan C, et al. Accepting critically ill transfer patients: adverse effect on a referral center's outcome and benchmark measures. *Ann Intern Med* 2003 Jun 3;138(11):882-90.

27 Seferian EG1, Afessa B, Gajic O, et al. Comparison of community and referral intensive care unit patients in a tertiary medical center: evidence for referral bias in the critically ill. Crit Care Med. 2008 Oct;36(10):2779-86.

28 Whisnant JP, Sacco SE, O'Fallon WM, et al. Referral bias in tertiary care: the aneurysmal subarachnoid haemorrhage. *J Neurosurg*. 1993; 78:726-32.

29 Rost NS, Bottle A, Lee JM, Randall M, et al. Global Comparators Stroke GOAL collaborators. Stroke Severity Is a Crucial Predictor of Outcome: An International Prospective Validation Study. *J Am Heart Assoc*. 2016; 21;5(1). pii: e002433. doi: 10.1161/JAHA.115.002433.

CHAPTER 6

The Dutch hospital standardised mortality ratio (HSMR) method and cardiac surgery: benchmarking in a national cohort using hospital administration data versus a clinical database

Pouw ME*, Siregar S*, Moons KGM, Versteegh MIM,
Bots ML, van der Graaf Y, Kalkman CJ, van Herwerden LA,
Groenwold RHH
* both authors contributed equally

Heart, 2014; 100(9):702-10

ABSTRACT

Objective

To compare the accuracy of data from hospital administration databases and a national clinical cardiac surgery database and to compare the performance of the Dutch Hospital Standardized Mortality Ratio (HSMR) method and the logistic EuroSCORE, for the purpose of benchmarking of mortality across hospitals.

Methods

Information on all patients undergoing cardiac surgery between 1 January 2007 and 31 December 2010 in 10 centres was extracted from The Netherlands Association for Cardio-Thoracic Surgery database and the Hospital Discharge Registry. The number of cardiac surgery interventions was compared between both databases. The EuroSCORE and HSMR models were updated in the study population and compared using the C-statistic, calibration plots and the Brier-score.

Results

The number of cardiac surgery interventions performed could not be assessed using the administrative database as the intervention code was incorrect in 1.4 to 26.3%, depending on the type of intervention. In 7.3% no intervention code was

registered. The updated administrative model was inferior to the updated clinical model with respect to discrimination (c-statistic of 0.77 versus 0.85, p-value <0.001) and calibration (Brier Score of 2.8% versus 2.6%, p-value <0.001, maximum score 3.0%). Two average performing hospitals according to the clinical model became outliers when benchmarking was performed using the administrative model.

Conclusions

In cardiac surgery, administrative data are less suitable than clinical data for the purpose of benchmarking. The use of either administrative or clinical risk-adjustment models can affect the outlier status of hospitals. Risk-adjustment models including procedure specific clinical risk factors are recommended.

INTRODUCTION

A valid comparison of outcomes between hospitals or healthcare providers (benchmarking) requires adjustment for severity of the health condition of patients and the performed interventions, often referred to as case-mix differences.[1-3] For this purpose prediction models have been developed to estimate risk-adjusted outcomes across hospitals. Most of these models are based on routinely collected administrative hospital data. For example, the hospital standardised mortality ratio (HSMR), first developed by Jarman in 1999 for the United Kingdom, is a risk-adjusted mortality rate calculated using prediction models based on administrative data.[4] Because administrative data are collected for other purposes, they are easily available, and thus the use of these data for benchmarking is cheap and requires relatively little extra effort.

However, administrative databases are often criticised for being inaccurate, incomplete, and containing limited information.[5-9] As a consequence, comparisons of risk-adjusted outcome rates between healthcare providers that are based on administrative database data might be unreliable, leading to unjustified criticism. For that reason clinical databases with corresponding clinical prediction models have been developed (e.g. European System for Cardiac Operative Risk Evaluation and Society of Thoracic Surgeons risk models in cardiac surgery) that include multiple clinical predictors for mortality.[10-12] The European System for Cardiac Operative Risk Evaluation (EuroSCORE) is a prediction model that was specifically designed to predict the risk of operative mortality related to cardiac surgery using 18 demographic and risk factors. The EuroSCORE can thus be used to adjust for differences in casemix in the comparison between healthcare providers. Models based on clinical risk factors are claimed to have a better predictive performance, resulting in improved risk-adjustment, and enable valid comparison of outcomes across centres.[5-7; 13] The downside is that clinical databases are more expensive; they comprise information that is obtained by active data collection by dedicated individuals and thus require continuous maintenance. Previous studies have not come to a conclusive answer to the question if clinical risk factors are necessary for adequate risk-adjustment. Some concluded that administrative data are sufficient to enable benchmarking, whereas others show a clear inferiority and insufficiency when compared to clinical data.[6-8; 13-20]

The aim of our study was to analyse whether a risk-adjustment model based on administrative data allows for adequate benchmarking in cardiac surgery. Using a nationwide cohort of cardiac surgery patients, we assessed the accuracy of an administrative database and the predictive performance of administrative models in comparison to a clinical database and the clinical EuroSCORE model.[21]

METHODS

Data

Both EuroSCORE and administrative variables of a national cohort of cardiac surgery patients in The Netherlands have been collected in two separate databases: 1.) The adult national cardiac surgery database of the Netherlands Association of Thoracic Surgery (NVT) and 2.) The National Hospital Discharge Registry (HDR) of The Netherlands.[21-23]

The adult national cardiac surgery database of the Netherlands Association of Thoracic Surgery This clinical database has a national coverage with participation of all sixteen centres performing cardiac surgery in The Netherlands.[21] All patients undergoing cardiac surgery excluding transfemoral aortic valve implantation, circulatory assist devices and pacemakers, are included in the database. Ten out of 16 cardiac centres participated in our study, in which 34229 consecutive procedures were performed between 1 January 2007 and 31 December 2010. Procedures with incomplete data were excluded (N=218, 0.6%), resulting in 34011 procedures for further analyses. The dataset consisted of predictors for mortality as listed in table 6.1, defined according to the EuroSCORE.[10] The EuroSCORE was developed to estimate the operative risk of mortality related to cardiac surgery (within 30-days and/or during the same hospital admission).[11] In this study, the EuroSCORE was used to estimate the risk of in-hospital mortality.

The Hospital Discharge Registry The Hospital Discharge Registry contains administrative data of all 10 hospitals included in this study. The dataset consists of patient characteristics and admission details such as age, comorbidity, sex and urgency of admission. For interventions the International Classifications of Health Interventions (ICHI) coding system is used and for diagnoses the International Classification of Disease (ICD-9).[24] The Dutch HSMR method is based on the HDR database and uses 50 risk-adjustment models, each for one specific group of diagnoses. The models estimate the risk of mortality for patients with a diagnosis belonging to the specific diagnose group.[22]

Linkage of datasets In order to compare the HDR and the NVT database and the models based on them, information on cardiac surgery interventions was required from both databases. Therefore, the HDR and NVT databases were linked to identify similar records. Both the HDR and NVT database contain anonymised data, meaning no directly identifying information is stored. Records from both databases were linked to the municipal registries based on date of birth, gender and zipcode, and were subsequently linked to each other. The linkage was performed by Statistics Netherlands and is described in previous publications.[21; 25; 26] The linkage of datasets is illustrated in the flowchart shown in figure 6.1. In total 26,178 (77%) records from the NVT database could be linked to a record in the HDR database and were used for further analyses. The predicted mortality according to the logistic EuroSCORE did not differ between the linked and the non-linked population (median 3.7%). Reasons for failed linkage were: the HDR record could not be linked to the municipal registries or no HDR record existed for the specific intervention (18.7%), the NVT record could not be linked to the municipal registries (2.7%), or no administrative model was available for the record (1.6%). The linkage of the HDR database to the municipal registries caused most linkage failure, as only 4 digits (out of 6) of the zipcode were available in the HDR database.

Comparison of data between the NVT and HDR database: intervention and in-hospital mortality

The type of intervention and the outcome in-hospital mortality were compared between the registries. Considering the fact that the NVT and the HDR registries use other risk factors for risk-adjustment, these were not compared. The NVT database was used as the reference for the type of intervention, because this information is collected by the surgeons themselves. The HDR database was used as the reference for in-hospital mortality, as the date of mortality is extracted directly from the up-to-date municipality registers. The comparison of in-hospital mortality between both databases was performed on patient level (as opposed to intervention level), to avoid persons being counted multiple times for mortality.

Comparison of risk-adjustment models

The administrative and clinical model – The Dutch Hospital Standardized Mortality Ratio (HSMR) method (models based on administrative data) and the logistic EuroSCORE (model based on clinical data) were applied in their original form to our study population, to predict the risk of in-hospital mortality in our study population.[10; 22] These models will subsequently be called

Administrative.1 and Clinical.1. Existing risk-adjustment models can be updated to a new study population. Updated models are adjusted to the characteristics of that population and are likely to show improved generalisability.[27] There are several methods to update a risk-adjustment model.[27] As cardiac surgery interventions are incorporated in multiple Dutch HSMR models (i.e. several diagnosis groups), one model for cardiac surgery was constructed using stepwise backward selection based on Akaike's Information Criterion.[28] This means that the intercept and the coefficients of all included covariates were re-estimated in our study population and only relevant risk factors were included in the updated model. To update the EuroSCORE model, the intercept and the coefficients of all included covariates were also re-estimated in our study population. This resulted in the updated models Administrative.2 and Clinical.2. The models can be updated even more thoroughly by inclusion of interaction terms, in order to maximise risk-adjustment in our study population. Thus, first-order interaction terms between all covariates were added to the updated models, resulting in the models Administrative.3 and Clinical.3.

Comparison of model performance The predictive performance of a risk-adjustment model is quantified by means of calibration and discrimination. Discrimination refers to the ability of a model to differentiate between subjects with and without the outcome and depends on the variables included in the model. The discrimination of the models was quantified using the area under the ROC-curve, which is equivalent to the c-statistic. The 95% CI of the c-statistic and the difference between two c-statistics was tested using DeLong's test.[29]

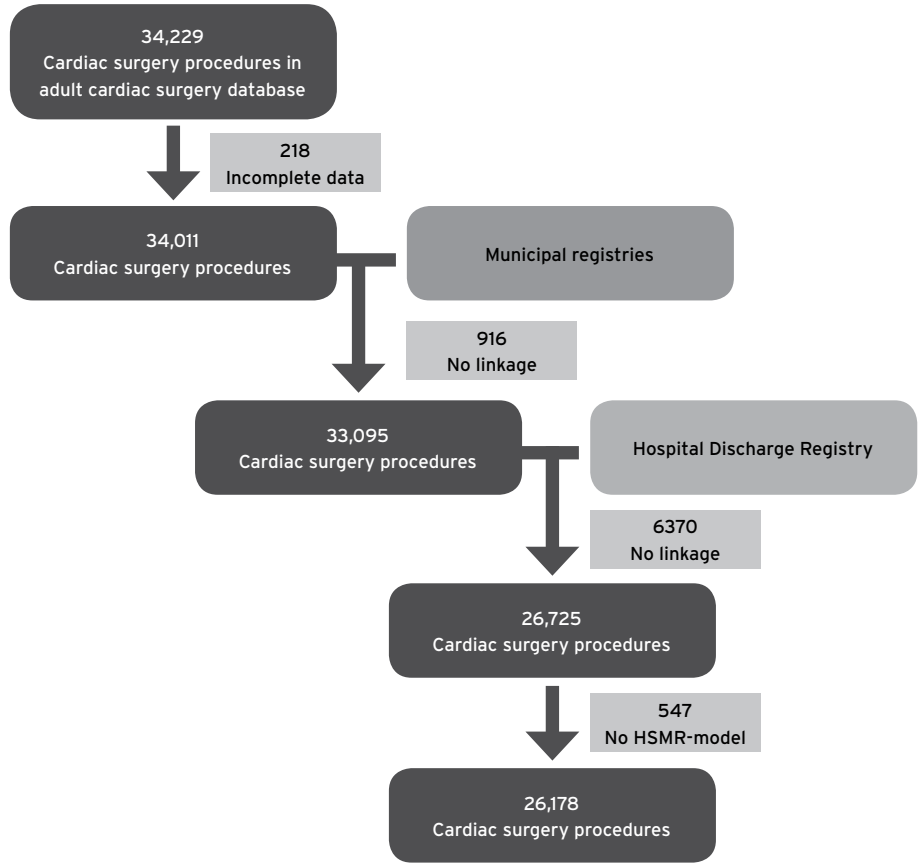
The calibration of a risk model refers to the ability of a model to predict how many patients will have the outcome. The calibration was assessed using calibration plots and the Brier Score. The Brier Score measures model accuracy on patient level by squaring and summing the difference between the predicted and the observed outcome per patient. The method by Redelmeier was used to estimate the 95% CI of the Brier Score and test the difference between two Brier scores.[30]

Benchmarking

In this study, benchmarking is performed by calculating the standardised mortality ratio (SMR) for all hospitals. The SMR is calculated by dividing the observed mortality with the expected mortality within a hospital. SMRs of the administrative and clinical models were compared. Centres with a SMR for which the 95% confidence intervals (CI) did not cover the value 1 were considered to be outliers. The 95% CI of the SMRs were estimated using the method described by Breslow and Day.[31]

All analyses were performed in R version 2.15.[32]

Figure 6.1 Flowchart of data flow. Data from the adult cardiac surgery database (Netherlands Association for Cardio-Thoracic Surgery) was linked to municipal registries and the Hospital Discharge Registry



RESULTS

Risk factor coding

The risk factors in the linked subset from both the administrative and clinical database are presented in table 6.1. Mean age was 66.6 years (+/- 10.7) and 29.5% of patients were female. A comparison of the prevalence of risk factors could not be made, as the definitions differed between the administrative and the clinical database.

Number of cardiac interventions performed (by type of intervention)

In total 14300 (54.6%) isolated CABG procedures were performed according to the NVT database. Other frequently performed interventions were: aortic valve replacement with or without concomitant CABG (12.1 and 8.3% respectively) and mitral valve repair with or without concomitant CABG (3.1 and 2.7% respectively). The proportion of isolated CABG, isolated aortic valve replacement, isolated mitral valve repair and isolated mitral valve replacement, which was coded with the correct main intervention code in the HDR ranged from 64.6% to 92.2% (table 6.2). The intervention code in the HDR was missing in 1923 (7.3%) procedures. As a result, the number of cardiac surgery interventions could not be accurately assessed using HDR data.

Calibration of the administrative models and the clinical models

Calibration of the risk models is shown in figure 6.2. Both the original models (Administrative.1 and Clinical.1) were poorly calibrated. Administrative.1 underestimated the risk of mortality, whereas Clinical.1 overestimated the risk of mortality. Updating improved calibration of both models, as the difference between observed and predicted mortality became smaller. However, in all model pairs the Brier Score for the administrative models remained significantly higher in comparison to the clinical models, indicating inferior calibration of the administrative model (table 6.3). The maximum Brier score in this data was 3.0%. Rescaling of the Brier Score on a scale from 0 to 100% would result in a score of 93.8% for Administrative.3 and 87.8% for Clinical.3.

Results were comparable in the subgroup analyses on isolated CABG procedures (figure 6.2 and table 6.3), where the maximum Brier score that was possible in these data was 1.3%.

Table 6.1 Variables recorded in the administrative database (HDR) and the clinical database (NVT)

Administrative variables	N (%) N = 26178	OR in updated model
Age <25 years (categories of 5 years up to >85)	66.5 (+ 10.7)	reference 0.29-1.01
Female Sex	7714 (29.5)	1.44***
Acute myocardial infarction	1899 (7.3)	1.25
Congestive heart failure	696 (2.7)	4.11***
Pulmonary disease	623 (2.4)	-
Renal disease	293 (1.1)	3.62***
Urgency	3292 (12.6)	2.20***
Peripheral vascular disease	551 (2.1)	2.17***
Cerebral vascular accident	241 (0.9)	2.75***
Peptic ulcer	51 (0.2)	4.36***
Social economic status		
Lowest	5450 (16.5%)	reference
Below average	5379 (16.3%)	0.89
Average	4999 (15.1%)	0.79*
Above average	5801 (17.5%)	0.70**
Highest	4541 (13.7%)	0.95
Unknown	6925 (20.9%)	
Year of discharge		
2007	6829 (20.6%)	reference
2008	6697 (20.2%)	1.04
2009	6941 (21.0%)	0.83
2010	5711 (17.3%)	0.69***
Unknown	6917 (20.9%)	
Admission from		
Home	19907 (60.2%)	reference
Nursing home	145 (0.4%)	3.41***
General hospital	4952 (15.0%)	1.27**
Academic centre	1174 (3.5%)	1.38*
Unknown	6917 (20.9%)	

Clinical variables	N (%) N = 26178	OR in updated model
Age (continuous)	66.6 (+ 10.7)	1.06***
Female sex	7714 (29.5)	1.33***
Recent myocardial infarction (<90 days)	3191 (12.2)	1.57***
LVEF 30–50%	4165 (15.9)	1.69***
LVEF <30%	1324 (5.1)	2.95***
Pulmonary disease	3019 (11.5)	1.79***
Serum creatinine >200 ↔mol/l	464 (1.8)	2.79***
Emergency operation	1317 (5.0)	2.38***
Extracardiac arteriopathy	3202 (12.2)	1.83***
Neurological dysfunction	780 (3.0)	1.26
Previous cardiac surgery	1709 (6.5)	2.78***
Systolic pulmonary pressure >60 mmHg	606 (2.3)	1.97***
Active endocarditis	216 (0.8)	1.45
Unstable angina	15776 (6.0)	1.95***
Critical preoperative state	983 (3.8)	2.51***
Ventricular septal rupture	47 (0.2)	3.93***
Other than isolated CABG	11809 (45.1)	3.43***
Thoracic aortic surgery	1258 (4.8)	2.75***

For dichotomous variables the number of patients and percentage of total population is reported; for continuous variables the mean and standard deviation. OR: odds ratio in multivariable logistic regression risk-adjustment model re-estimated in the study population; CABG: coronary artery bypass grafting; IQR: interquartile range; LVEF: left ventricular ejection fraction. *p < 0.05; **p < 0.01; ***p < 0.001.

Table 6.2 Comparison of intervention type and in-hospital mortality

			Hospital Discharge Registry (administrative data)		
	Intervention type		Correct main intervention code	Incorrect main intervention code	No code
NVT database (clinical data)	Isolated CABG	14300 (100%)	13185 (92.2%)	197 (1.4%)	918 (6.4%)
	Isolated AoV replacement	3157 (100%)	2461 (78.0%)	457 (14.5%)	239 (7.6%)
	Isolated MV repair	820 (100%)	625 (76.2%)	134 (16.3%)	61 (7.4%)
	Isolated MV replacement	316 (100%)	204 (64.6%)	83 (26.3%)	29 (9.2%)

NVT: Netherlands Association for Cardio-Thoracic Surgery, AoV: aortic valve, MV: mitral valve, CABG: coronary artery bypass grafting.

In-hospital mortality

In-hospital mortality in the HDR database is derived from the municipal registries and highly accurate. In the NVT database 42 of 762 (5.5%) patients who died during hospital stay were not coded as such and the other way around, 36 of 25005 (0.1%) survivors were incorrectly coded as in-hospital mortality during the same hospital admission.

Table 6.3 Brier Score of the three clinical models and the three administrative models, for all cardiac surgery and for only isolated coronary artery bypass surgery

	All cardiac surgery			Isolated CABG surgery		
Brier Scores	administrative	clinical	P value diff.	administrative	clinical	P value diff.
Original models	2.9 % [2.8-3.0]	3.0 % [2.8-3.2]	0.093	1.3 % [1.2-1.5]	1.4 % [1.1-1.7]	0.030
Updated models	2.9 % [2.7-3.1]	2.7 % [2.5-2.9]	<0.001	1.3% [1.1-1.4]	1.2 % [1.0-1.3]	<0.001
Updated + interaction terms	2.8 % [2.6-3.0]	2.6 % [2.5-2.8]	<0.001	1.2% [1.1-1.4]	1.2 % [1.0-1.3]	0.026

Brier scores range from 0 to a value depending on the prevalence of the outcome. The maximum Brier score that was possible in this data was 3.0% for all cardiac surgery and 1.3% for isolated CABG. A lower Brier score indicates better calibration. Brackets denote 95% confidence intervals. CABG: coronary artery bypass grafting.

Discrimination of the administrative models and the clinical models

Discrimination of the models is shown in figure 6.3. The c-statistics of the administrative models (0.756 – 0.788) are substantially lower than that of the clinical models (0.838 – 0.846), indicating inferior discrimination of the administrative models (p <0.001 for all three model pairs). Updating of the administrative model did not improve the discrimination (figure 6.3).

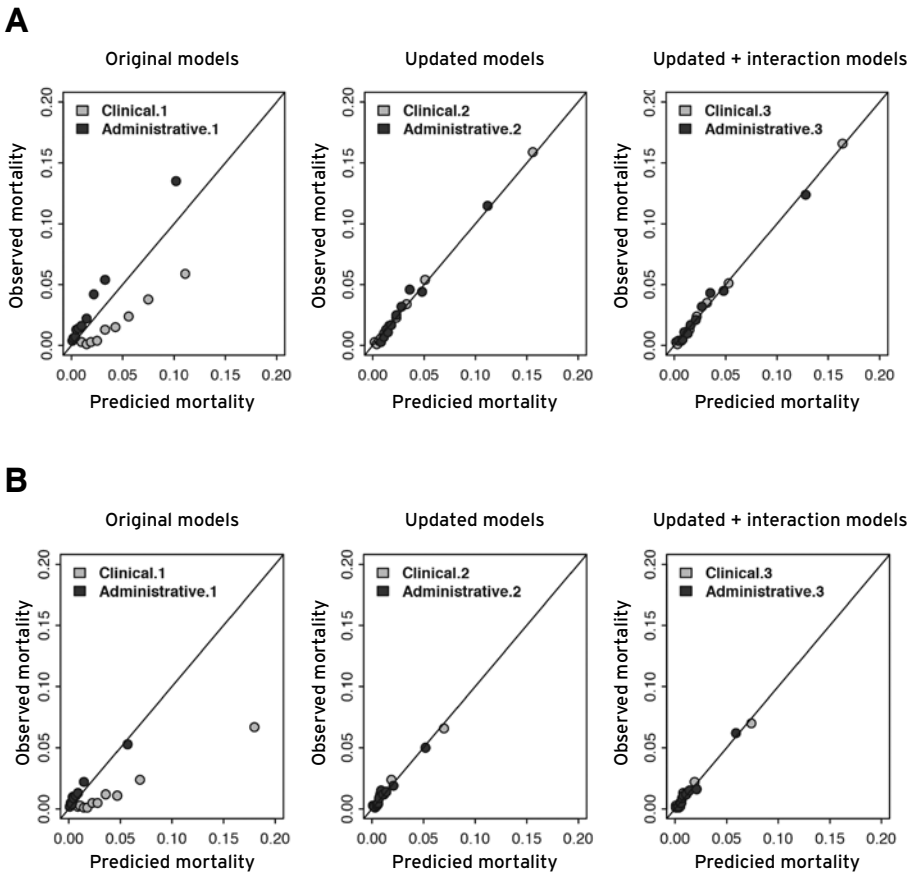
The effect on benchmarking

The effect of the use of administrative versus clinical models on benchmarking is shown in figure 6.4. The majority of SMR's calculated using the original administrative model were higher than 1, which indicates that the model underestimated the risk of mortality. For the original clinical model the opposite was found: the model overestimated the risk of mortality

Updating of models resulted in better predictions on hospital level (SMR's closer to 1). However, a considerable difference was found between the updated administrative versus the updated clinical models, for example in hospital B and hospital C (figure 6.4).The mean difference in SMR for admistrative.1 versus clinical.1 was 1.13 (range 0.23-2.08), 1.12 (range 0.004-0.37) for administrative.2 versus clinical.2, and 1.11 (range 0.001-0.43) for administrative.3 versus clinical.3.

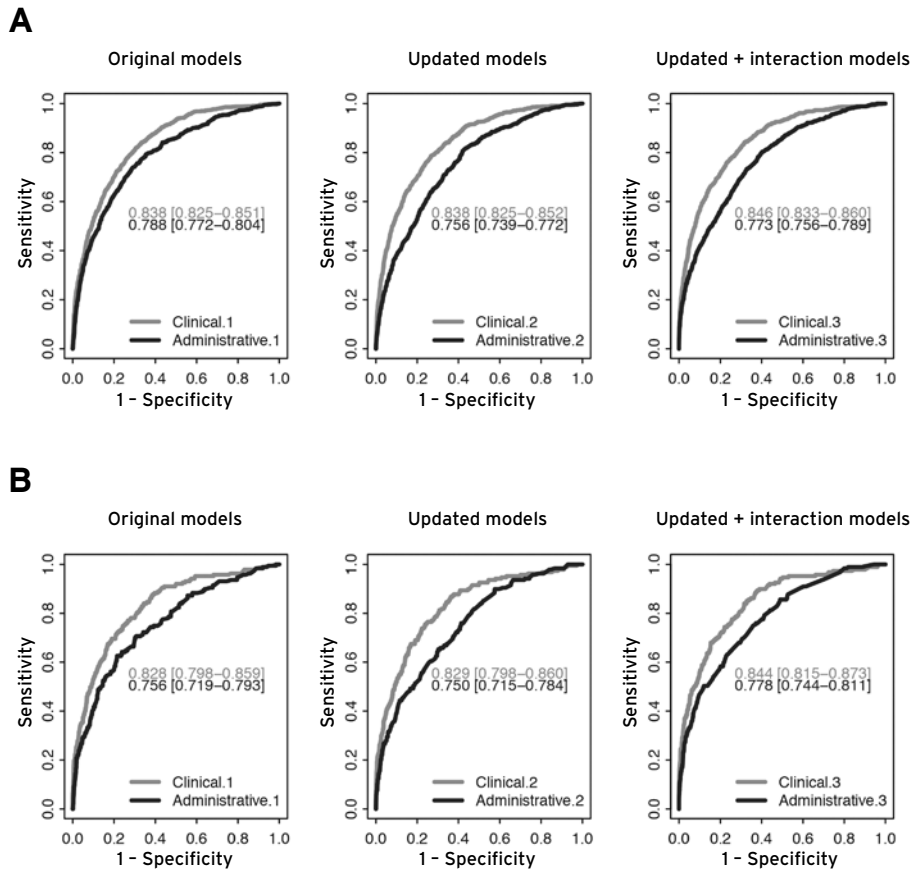
The SMR's calculated using the clinical and administrative models yielded different outliers. Hospital C and hospital J changed outlier status when either the updated model Administrative.3 or Clinical.3 were used. The analyses using only isolated CABG surgery yielded comparable results as those based on all cardiac surgery data (figure 6.4).

Figure 6.2 Calibration plot of the three clinical models and the three administrative models. The calibration plots of the clinical models are depicted in red and the calibration plots of the administrative models in blue



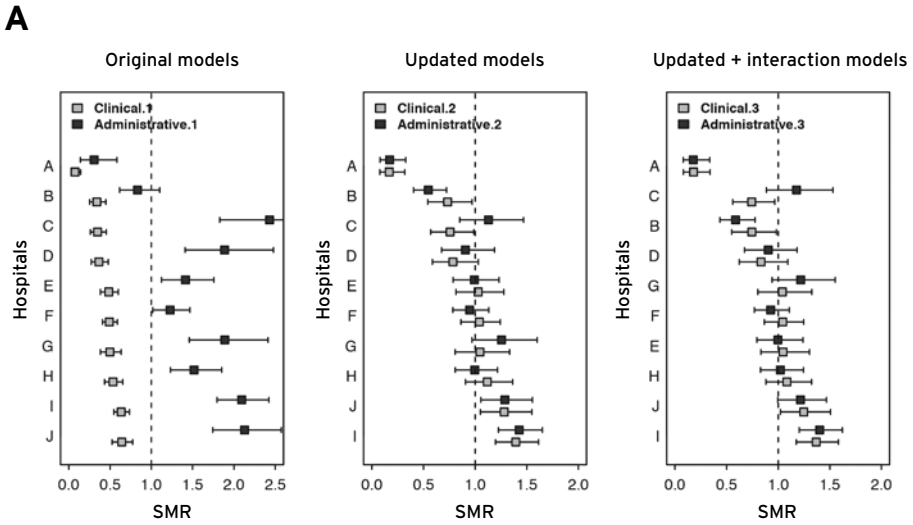
Panel A: models fitted on all cardiac surgery.
Panel B: models fitted on isolated coronary artery bypass grafting procedures.

Figure 6.3 Area under the ROC-curve of the clinical and the administrative models for the prediction of in-hospital mortality. The ROC-curves of the clinical models are depicted in red and the ROC-curves of the administrative models in blue



Panel A: models fitted on all cardiac surgery.
Panel B: models fitted on isolated coronary artery bypass grafting procedures.

Figure 6.4 Benchmarking using Standardized Mortality Ratio (SMR) calculated by the clinical models and the administrative models. SMRs of the clinical models are depicted in red and the SMRs of the administrative models in blue



Panel A: models fitted on all cardiac surgery.
Panel B: models fitted on isolated coronary artery bypass grafting procedures.

DISCUSSION

Principle findings

This study compared 1) data accuracy in the administrative HDR database to that in the clinical cardiac surgery database of the Netherlands Association of Cardio-Thoracic Surgery (NVT) and 2) the predictive performance of administrative models to the that of the clinical EuroSCORE model.

The reported intervention code in the administrative database was incorrect in up to 26%, depending on the type of surgery. As a result, the number of cardiac surgery interventions could not be accurately assessed.

After updating of the models to our data, the calibration of the administrative model was inferior to that of the clinical model. The importance of this shortcoming is marked by the identification of other outliers when used for benchmarking of hospitals.

Why models based on administrative data have inferior calibration and discrimination

When developing a risk prediction model, the first logical step is to consider which variables could be predictors for the outcome. However, administrative models are limited to the routinely collected variables, which might not necessarily be the strongest predictors. In our study, several strong predictors for mortality (shown in table 6.1) were not available in the administrative database. The other way around, administrative risk factors that were strongly associated with mortality had a low prevalence in our study population. This is likely to have affected the calibration and discrimination of the administrative models. Previous studies reported that much of the predictive performance of risk models is derived from a relatively small number of clinical variables and the predictive performance of administrative models could be improved with the addition of a limited number of clinical variables.[7; 13; 19; 33-34]

Why administrative data are inferior to clinical data for benchmarking purposes

The requirements of a risk-adjustment model depend on its goal. For benchmarking an adequate calibration is required: the model should adequately predict the expected mortality rate in a hospital. It can be seen as a scale that should weigh correctly. The performance of a scale mainly depends on its ability to weigh a (kilo)gram. If this feature is adequate, but the weighing is off par, the scale can be reset to zero to adjust it to any new situation. Similarly, the performance of a model depends on the strength of the predictors in the model (i.e. discrimination), as the model can be re-calibrated to update it in time or to make it suitable for a new population. It follows from aforementioned that the inferior discrimination of administrative models (in comparison to clinical models) will result in inferior calibration. It is shown in this study that this could very well affect the outlier status of a hospital.

Other issues in the use of administrative data

There are other reasons why the HDR database with routinely collected data turned out to be unsuitable for analyses of outcomes in cardiac surgery. Firstly, for a considerable number of records in our study population the intervention code was incorrect, unspecified (e.g. "cardiac surgery") or missing. Consequently, the number of cardiac surgery interventions performed could not be reliably assessed. Previous studies have also reported discrepant counts of operations in administrative data versus clinical data.[8; 17; 35]

Inaccurate coding could be attributed to the fact that data were collected by persons who were not actively involved in the clinical care and thus were dissociated from clinical information that could be necessary for correct reporting of data.[36] In addition, occasionally not all interventions and diagnoses are recorded. Also, admission and discharge dates are collected, instead of dates of intervention. This has been reported before as an important reason for variance in cardiac surgery volumes between administrative and clinical databases.[17]

Furthermore, the HSMR-method uses administrative models for specific diagnose codes. However, in cardiac surgery analyses of outcomes is performed by intervention type, as risk is considered to be mainly related to the performed intervention.

Implications for practice

The use of administrative data has many advantages over the use of clinical data. The data are routinely collected and stored, making them cheaper and readily available. However, the apparent benefits should be carefully weighed against the limitations and drawbacks of administrative models, when compared to clinical models.

Public benchmarking in general can be dangerous in the sense that the general public cannot be expected to understand the limitations and the prerequisites under which the results should be interpreted. The limitations are more pronounced for administrative data. This is particularly important because benchmarking could have far reaching consequences when known to healthcare consumers, the media, health insurance companies or governmental bodies. In this context, development of models with a high predictive performance, which might include clinical risk factors, should be strived for at all times.[37] If clinical data are already collected, their availability for benchmarking should be encouraged.

On the other hand, clinical data appeared to have an evident weakness as well. The outcome in-hospital mortality was misclassified in nearly 6% of the records in the clinical database used in this study. For outcomes such as vital state and readmissions, administrative databases were highly accurate, as information was derived from municipal registries. Administrative data sources could be used to verify outcomes data, thus complementing clinical databases. In this way, the strengths of both types of data are combined in order to optimise benchmarking in healthcare.[38]

The findings in this study are likely to hold true for populations other than cardiac surgery patients and in other countries in the world. Most probably, other specific surgical interventions such as for example oesophageal or hepato-biliary surgery, also require adjustment for risk factors not commonly included in administrative databases. Consequently, benchmarking in those populations will result in similar issues as encountered in this study.

Possible limitations

These analyses were based on data from 10 out of 16 cardiac surgery centres in The Netherlands. In general, the population of the six hospitals not participating in this study did not differ from the study population with regard to age, sex and the median logistic EuroSCORE. However, it is unknown if the results with regard to data accuracy are generalisable to all centres.

Secondly, the sensitivity of the linkage between the clinical and the administrative database was 77%. Although we did not find a difference in the overall risk profile between the linked and non-linked records, we do acknowledge that a substantial part of the total population were excluded from the analyses. We have no reason to believe that administrative models would perform any differently in the non-linked records or that data accuracy was better in the non-linked records. The conclusions of our study are thus unlikely to be affected by this limitation.

The goal of this study was to assess the accuracy of administrative data and the predictive performance of the accompanying models. As such, it was not our intention to design a new model for risk prediction in cardiac surgery. Thus, we chose to stay in line with the methods used to construct the original models and refrain from further sophisticated methods such as hierarchical modelling and shrinkage of coefficients.

The outcome in this study is in-hospital mortality. Several publications have previously shown why mortality at fixed time intervals is a more appropriate measure in outcomes evaluation. We acknowledge the limitations of this outcome and we are aware that mortality is one of the several indicators that can be used to measure quality, but certainly not the only one. For the purpose of our study, we have no reason to believe this has affected our results, as both the clinical and the administrative models were fitted on this outcome.

Conclusion

Although there are advantages to the use of administrative models for benchmarking in cardiac surgery, their calibration and discrimination (and thus performance in benchmarking) is inferior to that of clinical models. The use of either an administrative or a clinical model may affect the outlier status of hospitals. Therefore, in specific populations such as cardiac surgery, the use of prediction models including clinical risk factors is recommended.

REFERENCE LIST

1 Iezzoni LI. Risk adjustment for measuring healthcare outcomes. Ann Arbor, Mich: *Health Administration Press*; 1994.
2 Shahian DM, Normand SL, Torchiana DF, et al. Cardiac surgery report cards: comprehensive review and statistical critique. *Ann Thorac Surg* 2001;72:2155-68.
3 Heijink R, Koolman X, Pieter D, et al. Measuring and explaining mortality in Dutch hospitals; the hospital standardized mortality rate between 2003 and 2005. *BMC Health Serv Res* 2008;8:73.
4 Jarman B, Gault S, Alves B, et al. Explaining differences in English hospital death rates using routinely collected data. *BMJ* 1999;318:1515-20.
5 Bohensky MA, Jolley D, Pilcher DV, et al. Prognostic models based on administrative data alone inadequately predict the survival outcomes for critically ill patients at 180 days post-hospital discharge. *J Crit Care* 2012;27:422.e11-21.
6 Brinkman S, Abu-Hanna A, van der Veen A, et al. A comparison of the performance of a model based on administrative data and a model based on clinical data: effect of severity of illness on standardized mortality ratios of intensive care units. *Crit Care Med* 2012;40:373-8.
7 Hannan EL, Racz MJ, Jollis JG, et al. Using Medicare claims data to assess provider quality for CABG surgery: does it work well enough? *Health Serv Res* 1997;31:659-78.
8 Shahian DM, Silverstein T, Lovett AF, et al. Comparison of clinical and administrative data sources for hospital coronary artery bypass graft surgery report cards. *Circulation* 2007;115:1518-27.
9 Glance LG, Dick AW, Osler TM, et al. Accuracy of hospital report cards based on administrative data. *Health Serv Res* 2006;41:1413-37.
10 Nashef SA, Roques F, Michel P, et al. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg* 1999;16:9-13.
11 Roques F, Michel P, Goldstone AR, et al. The logistic EuroSCORE. *Eur Heart J* 2003;24:881-2.
12 Shahian DM, O'Brien SM, Filardo G, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 1--coronary artery bypass grafting surgery. *Ann Thorac Surg* 2009;88(1 Suppl):S2-22.
13 Geraci JM, Johnson ML, Gordon HS, et al. Mortality after cardiac bypass surgery: prediction from administrative versus clinical data. *Med Care* 2005;43:149-58.
14 Bratzler DW, Normand SL, Wang Y, et al. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. *PLoS One* 2011;6:e17401.
15 Gordon HS, Johnson ML, Wray NP, et al. Mortality after noncardiac surgery: prediction from administrative versus clinical data. *Med Care* 2005;43:159-67.
16 Hall BL, Hirbe M, Waterman B, et al. Comparison of mortality risk adjustment using a clinical data algorithm (American College of Surgeons National Surgical Quality Improvement Program) and an administrative data algorithm (Solucient) at the case level within a single institution. *J Am Coll Surg* 2007;205:767-77.
17 Mack MJ, Herbert M, Prince S, et al. Does reporting of coronary artery bypass grafting from administrative databases accurately reflect actual clinical outcomes? *J Thorac Cardiovasc Surg* 2005;129:1309-17.
18 Parker JP, Li Z, Damberg CL, et al. Administrative versus clinical data for coronary artery bypass graft surgery report cards: the view from California. *Med Care* 2006;44:687-95.
19 Ugolini C, Nobilio L. Risk adjustment for coronary artery bypass graft surgery: an administrative approach versus EuroSCORE. *Int J Qual Healthcare* 2004;16:157-64.
20 Hall BL, Hirbe M, Waterman B, et al. Comparison of mortality risk adjustment using a clinical data algorithm (American College of Surgeons National Surgical Quality Improvement Program) and an administrative data algorithm (Solucient) at the case level within a single institution. *J Am Coll Surg* 2007;205:767-77.
21 Siregar S, Groenwold RH, Versteegh MI, et al. Data Resource Profile: Adult cardiac surgery database of the Netherlands Association for Cardio-Thoracic Surgery. *Int J Epidemiology* 2013;42:142-149.

22 Jarman B, Pieter D, van der Veen AA, et al. The hospital standardised mortality ratio: a powerful tool for Dutch hospitals to assess their quality of care? *Qual Saf Healthcare* 2010;19:9-13.
23 Dutch Hospital Data. www.dutchhospitaldata.nl. Utrecht, The Netherlands.
24 World Health Organization. www.who.int. 2012.
25 Centraal Bureau voor de Statistiek. www.cbs.nl. Den Haag, The Netherlands.
26 Vaartjes I, Hoes AW, Reitsma JB, et al. Age- and gender-specific risk of death after first hospitalization for heart failure. *BMC Public Health* 2010;10:637.
27 Toll DB, Janssen KJ, Vergouwe Y, et al. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol* 2008;61:1085-94.
28 Steyerberg EW. Clinical prediction models. New York: Springer Science+Business Media; 2009.
29 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837-45.
30 Redelmeier DA, Bloch DA, Hickam DH. Assessing predictive accuracy: how to compare Brier scores. *J Clin Epidemiol* 1991;44:1141-6.
31 Breslow NE, Day NE. Rates and rate standardization. In: Heseltine E, editor. Statistical Methods in Cancer Research: Vol. II – The Design and Analysis of Cohort Studies. Lyon: International Agency for Research on Cancer; 1987. p. 48-80.
32 R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2011.
33 Jones RH, Hannan EL, Hammermeister KE, et al. Identification of preoperative variables needed for risk adjustment of short-term mortality after coronary artery bypass graft surgery. The Working Group Panel on the Cooperative CABG Database Project. *J Am Coll Cardiol* 1996;28:1478-87.
34 Simms AD, Reynolds S, Pieper K, et al. Evaluation of the NICE mini-GRACE risk scores for acute myocardial infarction using the Myocardial Ischaemia National Audit Project (MINAP) 2003-2009: National Institute for Cardiovascular Outcomes Research (NICOR). *Heart* 2013;99:35-40.
35 Aylin P, Bottle A, Majeed A. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *BMJ* 2007;334:1044.
36 Zhan C, Miller MR. Administrative data based patient safety research: a critical review. *Qual Saf Healthcare* 2003;12 Suppl 2:ii58-ii63.
37 Herrett E, Shah AD, Boggon R, et al. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. *BMJ* 2013;346:f2350.
38 Manda SO, Gale CP, Hall AS, et al. Statistical profiling of hospital performance using acute coronary syndrome mortality. *Cardiovasc J Afr* 2012;23:546-51.

CHAPTER 7

General discussion

GENERAL DISCUSSION

Hospital standardised mortality ratios (HSMRs) are published and used as a quality indicator in the Netherlands and many other countries. Because of several shortcomings in the current HSMR calculation, multiple authors have discouraged the use and publication of the HSMR.[1-5] However, HSMRs will most likely continue to be used as benchmark tool and performance measure by health decision makers, media and public. Because of the potential large impact of HSMR 'outlier status' on hospitals (reputation, financial situtation), the effects of these shortcomings must be studied and solutions or alternatives should be developed. In this way, some of the differences in HSMRs between hospitals can be put into perspective and explained. Furthermore, improvements of the HSMR calculation can be implemented on a nationwide scale to reduce the effect of these insufficiencies.

In this thesis we addressed several methodological challenges related to the HSMR (overall hospital standardised mortality ratio) and the SMR (diagnosis specific standardised mortality ratio). We investigated the model underlying the HSMR, the effects of adjustments to this model with respect to comorbidity and outcome, and explored whether looking into clinical subgroups alleviated some deficiencies of the HSMR.

Current shortcomings of the HSMR

Based on the studies elaborated in the previous chapters, it can be stated that the use of the HSMR in its current form as a performance indicator is subject to various forms of bias, and therefore less useful for hospital profiling or comparison. The HSMR compares the number of in-hospital deaths of a hospital with the expected number of deaths for that hospital as computed with a casemix correction model. Unfortunately, the current casemix correction model is far from perfect.

First, to account for severity of illness of admissions, the HSMR casemix adjustment model includes the Charlson Comorbidity measure to reflect patients' existing comorbidities, ideally only those present on admission. However, in most administrative databases, used as the basis for HSMR calculation, comorbidities of patients are difficult to distinguish from complications that occurred during hospitalisation. For obvious reasons, the casemix adjustment model should not incorporate complications during hospitalisation, as these could be the result from medical misadventures or indeed poor quality of care itself – hence adjusting for these complications could lead to overestimation

of expected mortality. The Elixhauser comorbidity measure tries to differentiate complications from comorbidities, hence calculating HSMR based on the Elixhauser measure seems fairer. In chapter 3 we compared the HSMR calculation based on the Elixhauser measure with the Charlson measure and found that HSMR models estimated with the Charlson measure and the Elixhauser measure perform practically similarly in terms of discrimination. However, the performance of Elixhauser HSMR models may be underestimated in our analysis, as hospitals currently pay specific attention to registration of comorbidities included in the Charlson measure, neglecting comorbidities registered in the Elixhauser measure.

Secondly, the current HSMR model is vulnerable to discharge and referral bias as it only looks into mortality occurring during admission. In chapter 4 we uncovered evidence for discharge and referral bias when an in-hospital timeframe is used to calculate HSMRs, while 30 days post-admission and 30 days post-discharge timeframes are less affected by differences in discharge practices.

Thirdly, in chapter 5 we discuss that some aspects of referral bias are still not taken into account, even when including post-discharge mortality in the calculation of HSMRs. The lack of distinction between high-risk 'step-up' and low-risk 'step-down' referrals within prediction models ignores any potential difference in expected mortality between high-risk and low-risk referrals. Especially HSMRs of specialised hospitals may be affected, as these hospitals tend to receive more 'step-up/high risk' referrals than other hospitals.

Fourthly, the lack of important (clinical) variables in the models used to compare the quality of care between hospitals, can also cause casemix bias.[6-9] In chapter 5 we found that according to SMR calculations for two specific neurological disorders – traumatic brain injury and subarachnoid haemorrhage – specialised neurologic, tertiary care centers appeared to perform worse than non-specialised, regional hospitals. This is counter intuitive and contrasts with other studies where specialised centers were found to perform better than non-specialised hospitals.[10-12] However, all these preceding studies developed their own SMR prediction models and included many important clinical variables in their models. Such variables were lacking in the administrative database that was used for our analysis in chapter 5. Indeed, in Chapter 6 we report that for a specific patient group undergoing cardiac surgery, casemix correction models based on data that did contain relevant clinical variables, had higher calibration and discrimination than models based on administrative data usually lacking such detailed clinical variables.

Finally, it is important to realise that even with a 'perfect' casemix adjustment model, systematic bias can occur if the distributions of casemix between

hospitals are not *proportionally* the same. This is a consequence of the indirect standardisation method that is used in the HSMR calculation (chapter 2). Due to this so-called indirect standardisation method, HSMRs may not automatically be comparable, neither across hospitals nor for a single hospital over time, unless the underlying casemix distributions are proportionally the same or when there is no interaction between hospital and casemix. In chapter 2 we found evidence of interaction between hospitals and the casemix variables 'urgency of admission' and 'comorbidity'.

Recommendations for adjustment and use of HSMR

Given that the HSMR is likely here to stay and based on the findings in this thesis we propose a series of recommendations. Table 7.1 summarizes the recommendations, the specific shortcomings addressed by each of these recommendations, and the related chapter(s) in this thesis.

Table 7.1 List of recommendations

Recommendation	Reduces problem of:	Chapter
Strive for enhancement of the casemix model by the inclusion of:		
1. Elixhauser comorbidity measure instead of Charlson comorbidity measure	Casemix bias	3
2. Survival status (death/alive) shortly after discharge	Referral and discharge bias	4, 5
3. Reason of referral (step up/down)	Referral bias	5
4. Relevant clinical predictors	Casemix bias	5, 6
Zoom in onto comparable entities by:		
5. Clustering hospitals with approximately the same casemix, and only compare SMRs of hospitals within such a cluster	Indirect standardisation Referral bias	2, 5
6. Using diagnostic specific SMRs instead of overall HSMR	Indirect standardisation	2, 5, 6

Strive for enhancement of the casemix model

The first set of recommendations concerns the improvement of the casemix correction model used to calculate standardised mortality ratios.

First, we prefer using the Elixhauser comorbidity measure to the Charlson comorbidity measure in the HSMR calculation. The Elixhauser comorbidity measure excludes complications, which is not the case with the Charlson comorbidity measure. Therefore – on theoretical grounds (described in chapter 3) – the Elixhauser comorbidity measure reduces the chance of casemix bias in the HSMR model.

Secondly, post-discharge mortality should be used as the outcome in the HSMR model to reduce discharge and referral bias (chapter 4). Of course, mortality after discharge is also affected by factors beyond hospital control, such as quality of out-patient care or quality of other referring and admitting hospitals. From a societal perspective this could actually be beneficial, as hospitals now have a stake in organizing optimal handover and post-discharge care.

Thirdly, additional information about the type of referral should be included in the HSMR model, such that the model is able to differentiate between high-risk referrals and low-risk referrals (chapter 5).

Fourthly, we recommend to investigate whether inclusion of additional clinical variables in the casemix correction model of a specific diagnosis is required. For diagnoses such as traumatic brain injury, subarachnoid haemorrhage or patients undergoing cardiac surgery, important (clinical) predictors of mortality are not routinely collected in administrative databases and therefore currently not included in the HSMR model (see chapter 5 and 6). As there may be clinical diagnoses for which the burden of collecting additional clinical variables does not outweigh the gain in reducing casemix bias, this needs to be examined for each clinical diagnosis separately.

Zoom in onto comparable entities

The second set of recommendations concerns the method of comparing and reporting standardised mortality ratios to make the numbers more informative and therefore more usable for hospital profiling.

We recommend to investigate whether certain types of hospitals have a deviant casemix (and different disease severity) for the diagnosis under investigation and to determine whether these groups of hospitals should be examined separately (chapter 2, 5). Benchmarking hospitals with comparable

hospitals is essential, especially when information about (known) relevant clinical predictors is lacking in the administrative database and, as a consequence, casemix correction is suboptimal. For example, SMRs of traumatic brain injury of specialised neurosurgical, tertiary care centers – without information about the Glasgow Coma Scale or Injury Severity Score at admission – could be cautiously compared with other tertiary care centers but definitely not with non-specialised, regional hospitals (chapter 5).

As hospitals are usually not specialised in every diagnosis, clustering and benchmarking hospitals should depend on the diagnosis under investigation. For example, a hospital specialised in neurological diseases may not be a cardiac centre. So, for neurological diseases the hospital could be compared with other hospitals specialised in neurological diseases but for cardiac diseases it should not be compared with cardiac centres.

Therefore, in conjunction with the previous recommendation, we recommend to report diagnosis-specific SMRs, instead of an overall hospital standardised mortality ratio. In this way, bias resulting from indirect standardisation is reduced (chapter 2). More importantly, the numbers will be more informative for various stakeholders such as patients, doctors, and health authorities. For example, a patient with a cardiovascular disease is less interested in the overall performance of a hospital but all the more interested in the performance of the departments of cardiology / cardiosurgery. Also, a cardiologist can now compare the performance of his department with equivalent cardiology departments of other hospitals.

An 'average' overall HSMR has no informative value as the SMR of the diagnosis of interest may be 'diluted' by SMRs of other diagnoses included in the overall HSMR (chapter 2). Studies have shown that differences in performance of care *within* hospitals (across diagnosis groups), are at least as large as differences *between* hospitals.[13, 14] If negative (media) attention is drawn towards a hospital this is most often because a specific department in that hospital is underperforming, and not because of overall hospital performance.[15-19] To take this one step further, using a more selective approach to utilizing SMRs at the level of a common diagnosis or frequently performed procedure, the quality of an entire hospital can be judged by observing the number of high-performing departments. To define high-performance, not only mortality but also patient-centred outcomes and patient satisfaction with delivered care can be taken into account. A large number of high performing departments most likely reflects an overall safety and quality conscious hospital culture. Conversely, a large number of underperforming departments might indicate governance and coordination problems and a poor safety culture at the hospital level.

Future perspectives

The proposed recommendations for adjusting the HSMR will reduce several current shortcomings of the HSMR. However, there are other limitations in using the HSMR that need to be addressed in future studies.[20-25]

One of the shortcomings is that HSMR is subject to bias because of constant risk fallacy – occurring if the association between casemix and outcome (mortality) differs between hospitals.[23, 26] One example of this is when patients are admitted more often for the same (chronic) disease, for instance oncologic patients who are admitted several times for each new round of chemotherapy.[20] If they decease during such an admission, it can only be during the last one, resulting in many admissions with hospital survival as the outcome. If admission patterns for such patients differ between hospitals, this may lead to bias.

Another shortcoming is the coding variation between hospitals.[24, 25, 27] In addition, data on comorbidity are frequently missing and contain errors.[25] Effort should be made – e.g. by allocating resources for coding and education of encoders – to improve registration and coding uniformity. Well-designed hospital electronic medical records with intuitive user interfaces should make it extremely easy for doctors and nurses to correctly code primary and secondary diagnoses, outcomes and complications with minimal effort.

Finally, a central problematic issue regarding HSMRs is whether mortality truly reflects quality of care.[2-4] The degree of delivered quality of care is not always directly related with mortality, e.g. cataract operations of suboptimal quality might result in all sorts of complications such as infections or blindness but rarely to death. Another example is the fact that the relationship of the standardised mortality ratio with other measures of quality is considered weak and inconsistent.[4]

High-quality of care has been defined as care that is safe, effective, patient-centred, timely, efficient and equitable.[28] Mortality rates are assumed to reflect only a few of these aspects. As care consists of different aspects, different stakeholders (patients, doctors, patient safety agencies, administrators) might also prioritise these aspects differently. For example, doctors might value office hours, waiting times at the clinic, waiting times for the communication of test results to patient, parking spaces, and quality of food less than patients. Creating a dashboard of performance indicators for hospitals might cover more aspects and interests of hospitals' performance.

Conclusion

Considering the current shortcomings of the HSMR, opponents of the HSMR argue that this measure will do more harm than good.[1-5] Therefore, they recommend to *not* use HSMRs for hospital profiling. However, not publishing mortality rates immediately insinuates that a hospital may have something to hide. Moreover, for some high-risk procedures that are heavily dependent on specialised skills and a well-organized care pathway – for example, heart surgery – SMRs are indeed an important indicator of quality of care.[29]

Therefore, in conclusion, we advocate to use (and publish) diagnosis specific SMRs using a casemix correction model with (1) the Elixhauser comorbidity measure, (2) additional diagnosis specific relevant (clinical) predictors if necessary, and (3) early post-discharge mortality as an outcome. Furthermore, as long as relevant predictors are not available for the concerning diagnosis, identifying clusters of hospitals with approximately the same casemix and treatment options is pivotal to use SMRs as benchmark tool. Extreme caution is still needed with the interpretation of SMR numbers and information concerning the pitfalls and shortcomings should be actively communicated to all stakeholders.

REFERENCE LIST

1 Black N. Assessing the quality of hospitals. *BMJ* 2010; 340:c2066
2 Lilford R, Pronovost P. Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *BMJ* 2010; 340:c2016.
3 Goodacre S, Campbell M, Carter A. What do hospital mortality rates tell us about quality of care? *Emerg Med J* 2015; 32:244-7
4 Pitches DW, Mohammed MA, Lilford RJ. What is the empirical evidence that hospitals with higher-risk adjusted mortality rates provide poorer quality care? A systematic review of the literature. *BMC Health Serv Res* 2007; 7:91
5 Van Gestel YRBM, Lemmens VEPP, Lingsma HF, et al. The Hospital Standardized Morality Ratio Fallacy. A Narrative Review. *Med Care* 2012; 50:662-7
6 Katzan IL, SPertus J, Bettger JP, et al. Risk adjustment of ischemic stroke outcomes for comparing hospital performance. *Stroke* 2014 doi: 10.1161/01.str.0000441948.35804.77.
7 Brinkman S, Abu-Hanna A , van der Veen A, et al. A comparison of the performance of a model based on administrative data and a model based on clinical data: effect of severity of illness on standardized mortality ratios of intensive care units. *Crit Care Med* 2012; 40:373-8 .
8 Bohensky MA, Jolley D, Pilcher DV, et al. Prognostic models based on administrative data alone inadequately predict the survival outcomes for critically ill patients at 180 days post-hospital discharge. *J Crit Care* 2012; doi: 10.1016/j.jcrc.2012.03.008.
9 Shahian DM, Silverstein T, Lovett AF, et al. Comparison of clinical and administrative data sources for hospital coronary artery bypass grafts surgery report cards. *Circulation* 2007; 115:1518-27.
10 Matsui H, Fushimi K, Yasunaga H. Variation in risk-standardized morality of stroke among hospitals in Japan *PloS ONE* 2015; doi: 10.1371/journal.pone.0139216.
11 Saposnik G, Baibergenova A, O'Donnell M, Hill MD, Kapral MK, Hachinski V, et al. Hospital volume and stroke outcome: does it matter? *Neurology* 2007; 69: 1142–1151. PMID: 17634420.
12 Spijkers ATE, Meylaerts SAG, Leenen LPH. Mortality decreases by implementing a level 1 trauma center in a Dutch hospital. *J Trauma* 2010; 69:1138-42. doi: 10.1097/TA.0b013e3181e12526.
13 Jha AK, Li Z, Orav EJ, et al. Care in U.S. hospitals – the Hospital Quality Alliance program. *N Engl J Med* 2005; 353:265-74.
14 Wilson B, Thornton JG, Hewison J, et al. The Leeds University Maternity Audit Project. *Int J Qual Health Care* 2002; 14:175-81.
15 https://www.onderzoeksraad.nl/uploads/items-docs/364/rapport_hartchirurgie_sint_radboud.pdf
16 http://www.nrc.nl/handelsblad/2005/10/17/commotie-over-verhoogd-risico-radboud-ziekenhuis-11029965
17 http://www.nu.nl/binnenland/4158726/umc-utrecht-verzwijgt-twee-dodelijke-incidenten-kno-afdeling.html
18 http://nos.nl/artikel/2099107-umc-utrecht-onder-verscherpt-toezicht-na-dodelijke-incidenten.html
19 http://nos.nl/artikel/2090110-kno-afdeling-umc-utrecht-weer-in-opspraak.html
20 van den Bosch WF, Kelder JC, Wagner C. Predicting hospital mortality among frequently readmitted patients: HSMR biased by readmission. *BMC Health Serv Res* 2011; 11:57.
21 Penfold RB, Dean S, Flemons W, et al. Do hospital standardized mortality ratios measure patient safety? HSMRs in the Winnipeg Regional Health Authority. *Healthc Pap.* 2008; 8:8-24; discussion 69-75.
22 Cohen J, Bilsen J, Addington-Hall J, et al. Population-based study of dying in hospital in six European countries. *Palliat Med* 2008; 22:702-10
23 Mohammed MA, Deeks JJ, GIrling A, et al. Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals. *BMJ* 2009; 338:b780.
24 Van den Bosch WF, Silberbusch J, roozanedaal KJ, et al. Variations in patient data coding affect hospital standardized mortality ratio (HSMR). *Ned Tijdschr Geneeskd* 2010; 154:A1189
25 Kalkman CJ, Rinkel GBA, Roos Y. Coderen van codes nog veel te divers. *Medisch Contact* 2013; 9:476-8
26 Debray TPA, Vergouwe Y, Koffijberg H, et al. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol* 2015; 68:279-89
27 Shojania KG, Forster AJ. Hospital mortality: when failure is not a good measure of success. Can Med Assoc J 2008; 179:153-7
28 Institute of Medicine (IOM). Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, D.C: *National Academy Press*; 2001.
29 Ettinger WH, Hylka SM, Phillips RA, et al. When things go wrong: the impact of being a statistical outlier in publicly reported coronary artery bypass graft surgery mortality data. *Am J Med Qual* 2008; 23:90-5.

CHAPTER 8

Summary in Dutch
Samenvatting

Acknowledgements in Dutch
Dankwoord

Curriculum vitae

SUMMARY IN DUTCH
SAMENVATTING

Ziekenhuissterfte wordt beschouwd als één van de belangrijkste indicatoren voor de kwaliteit van zorg van een ziekenhuis. Het vergelijken van sterftecijfers tussen ziekenhuizen is echter gecompliceerd, aangezien sterfte beïnvloed wordt door factoren die niet perse gerelateerd zijn aan kwaliteit van zorg. Eén van de belangrijkste verklaringen voor verschillen in sterfte zijn de verschillen tussen ziekenhuispopulaties, de zogenaamde 'casemix' verschillen. Het gestandaardiseerde sterftecijfer (HSMR) beoogt rekening te houden met deze casemix verschillen zodat ziekenhuizen onderling vergeleken kunnen worden. De HSMR vergelijkt de *werkelijke* ziekenhuissterfte met een verwachte ziekenhuissterfte. De *verwachte* ziekenhuissterfte wordt berekend met behulp van een 'casemix correctie' model. De assumptie is dat een hoger dan verwachte ziekenhuissterfte 'vermijdbare' sterfte weerspiegelt en dus een lagere kwaliteit van zorg.

De HSMR wordt in veel landen, waaronder in Nederland, jaarlijks gepubliceerd en gebruikt als kwaliteitsindicator. Sinds de introductie van de HSMR zijn zorgen geuit over de tekortkomingen met betrekking tot de methode van berekening en de betrouwbaarheid als kwaliteitsindicator. Aangezien publicatie van de HSMR verstrekkende gevolgen kan hebben voor de reputatie van een ziekenhuis, is het van groot belang dat de gevolgen van deze tekortkomingen bestudeerd worden.

Dit proefschrift richt zich op verscheidene methodologische aspecten van de HSMR (ziekenhuis breed gestandaardiseerde sterftecijfer) en de SMR (diagnose specifiek gestandaardiseerde sterftecijfer). Het onderliggende casemix correctie model wordt bestudeerd alsmede de effecten van aanpassingen aan dit model.

Huidige tekortkomingen van de HSMR

De huidige HSMR en het onderliggende casemix correctie model zijn onderhevig aan verschillende vormen van bias en dus suboptimaal.

Het HSMR casemix model includeert comorbiditeiten van de patiënt om de verwachte sterfte te berekenen. Hierbij dient het model alleen rekening te houden met comorbiditeiten bij opname en niet met complicaties die optreden tijdens de ziekenhuis opname; complicaties kunnen immers het resultaat zijn van slechte kwaliteit van zorg. In het huidige model worden comorbiditeiten beschreven aan de hand van de Charlson Comorbidity Index. Het

blijkt echter lastig om comorbiditeiten en complicaties van elkaar te onderscheiden in de administratieve databases.

Daarom heeft Elixhauser et al. een comorbiditeitsmaat ontwikkeld waarbij getracht wordt onderscheid te maken tussen complicaties en comorbiditeiten. In hoofdstuk 3 wordt het casemix correctie model berekend met de Elixhauser comorbiditeiten vergeleken met het model berekend met de Charlson comorbiditeiten. Aangezien het discriminerend vermogen van de twee type modellen in ons onderzoek nagenoeg identiek is, zou op basis van theoretische gronden, casemix correctie modellen berekend met de Elixhauser comorbiditeiten de voorkeur moeten hebben.

Ten tweede is het huidige HSMR model gevoelig voor vertekening door verschillen in ontslag- en verwijsbeleid van ziekenhuizen, zogeheten ontslag- en verwijsbias, omdat alleen sterfte tijdens ziekenhuisopname wordt meegenomen. In hoofdstuk 4 wordt beschreven hoe de HSMR verandert wanneer sterfte tot 30 dagen na opname of tot 30 dagen na ontslag wordt meegenomen. Deze maten zijn minder gevoelig voor verschillen tussen ziekenhuizen in ontslagbeleid en verwijsbeleid.

Ten derde kan verwijsbias optreden (hoofdstuk 5) indien het casemix correctie model geen onderscheid maakt tussen vanuit elders verwezen patiënten met een hoog sterfterisico versus verwezen patiënten met een laag sterfterisico. Qua karakteristieken, en dus qua casemix, worden deze patiënten dan als gelijk beschouwd; echter hun verwachte mortaliteit is wel degelijk verschillend. Hierdoor zullen HSMRs van gespecialiseerde ziekenhuizen, zoals academische ziekenhuizen, hoger zijn dan mag worden verwacht, omdat deze ziekenhuizen relatief meer hoog-risico verwijzingen ontvangen dan andere.

Ten vierde kan casemix bias optreden indien belangrijke (klinische) voorspellers van sterfte ontbreken in de HSMR berekeningen. Hoofdstuk 5 beschrijft het contra-intuïtieve resultaat waarbij ziekenhuizen gespecialiseerd in de neurologische aandoeningen 'traumatisch hersenschade' en 'subarachnoïdale bloedingen' slechtere SMR scores hebben dan niet-gespecialiseerde ziekenhuizen. Gespecialiseerde neurochirurgische ziekenhuizen zullen over het algemeen ziekere patiënten met deze neurologische aandoeningen opnemen dan niet-gespecialiseerde neurochirurgische ziekenhuizen. Indien deze verschillende type ziekenhuizen met elkaar vergeleken worden, moet rekening gehouden worden met de ernst van de ziekte van de patiënten. Hiervoor worden predictoren, die relevant zijn voor het voorspellen van mortaliteit bij deze twee neurologische aandoeningen, gebruikt (bijvoorbeeld de Glasgow Coma Scale en de Injury Severity Score). In onze berekeningen zijn deze variabelen niet meegenomen omdat ze niet worden bijgehouden in de gebruikte administratieve database. In hoofdstuk 6 bestuderen we wat het effect is als

belangrijke (klinisch) relevante variabelen wel meegenomen worden in de casemix correctiemodellen. Casemix correctiemodellen voor hartchirurgische patiënten gebaseerd op een database met klinisch relevante variabelen hebben hogere kalibratie en discriminatie dan modellen gebaseerd op administratieve databases zonder deze relevante variabelen.

Tenslotte is het belangrijk om te realiseren dat de HSMR altijd onderhevig is aan systemische bias als de casemix distributies van ziekenhuizen niet proportioneel gelijk zijn. Dit komt doordat de HSMR berekent wordt met de indirecte standaardisatie methodiek. In hoofdstuk 2 wordt de indirecte standaardisatie methodiek onder de loep genomen en voorbeelden en consequenties van deze methodiek op de HSMR bestudeerd.

Aanbevelingen

Op basis van onze bevindingen in dit proefschrift doen we een aantal aanbevelingen. In tabel 8.1 staan de aanbevelingen, de bijbehorende tekortkoming en het hoofdstuk vermeld.

Tabel 8.1 Lijst met aanbevelingen

Aanbeveling	Vermindert probleem van:	Hoofdstuk
Streef naar verbetering van de casemix model door inclusie van:		
1. Elixhauser comorbiditeit maat in plaats van Charlson comorbiditeit maat	Casemix bias	3
2. Survival status (dood/leven) kort na ontslag	Verwijs- en ontslagbias	4,5
3. Type verwijzing (hoog/laag risico)	Verwijs bias	5
4. Relevante klinische variabelen	Casemix bias	5, 6
Vergroot vergelijkbaarheid van ziekenhuizen door:		
5. Clusteren van ziekenhuizen met nagenoeg dezelfde casemix	Indirecte standaardisatie Verwijs bias	2,5
6. Gebruiken van diagnose specifieke SMRs in plaats van HSMRs	Indirecte standaardisatie	2,5,6

Streven naar verbeteringen van het casemix model

De eerste set van aanbevelingen betreft de verbetering van het casemix correctie model van de HSMR.

Ten eerste bevelen we het gebruik van de Elixhauser comorbiditeit maat aan in plaats van de Charlson comorbiditeit maat aangezien de Elixhauser comorbiditeit maat onderscheid maakt tussen complicaties en comorbiditeiten. Theoretisch gezien vermindert de Elixhauser comorbiditeit maat hierdoor de kans op casemix bias in het HSMR model (hoofdstuk 3).

Ten tweede moet ook sterfte kort na ontslag meegenomen worden in het HSMR model zodat er minder kans op verwijs- en ontslagbias is (hoofdstuk 4 en 5).

Ten derde is informatie nodig over de 'fase van ziekte' van een doorverwezen patiënt zodat onderscheid gemaakt kan worden tussen hoog-risico en laag-risico verwijzingen (hoofdstuk 5).

Tenslotte, bevelen we aan om voor elke diagnose te onderzoeken of er nog ontbrekende (klinische) predictoren nodig zijn om de casemix correctie model te optimaliseren (hoofdstuk 5 en 6).

Focus op vergelijkbare ziekenhuizen

De tweede set van aanbevelingen betreft de methode om sterftecijfers te vergelijken en te rapporteren.

Allereerst moeten ziekenhuizen met een afwijkende casemix apart bestudeerd worden. Dit is met name van belang als op voorhand reeds bekend is dat het casemix correctie model niet optimaal is, bijvoorbeeld door het ontbreken van belangrijke klinische variabelen (hoofdstuk 2 en 5).

Vervolgens moeten, in samenhang met de eerste aanbeveling, diagnose specifiek SMRs gerapporteerd worden. Omdat ziekenhuizen niet gespecialiseerd kunnen zijn in alle diagnoses, moet het clusteren van ziekenhuizen gebeuren op basis van de specifieke, te analyseren diagnose. Bijvoorbeeld een neurochirurgisch centrum hoeft nog geen cardiochirurgisch ziekenhuis te zijn. Voor neurologische diagnoses moet het ziekenhuis vergeleken worden met andere neurochirurgische centra terwijl het voor hartaandoeningen niet met gespecialiseerde cardiochirurgische centra vergeleken moet worden.

Op deze wijze zijn de cijfers informatiever voor de verschillende belanghebbenden zoals patiënten, artsen en zorgautoriteiten.

Conclusie

In dit proefschrift zijn verscheidene methodologische gebreken van de HSMR geïdentificeerd. Daarnaast zijn er aanbevelingen gedaan om de HSMR methodologisch te verbeteren.

Concluderend, pleiten we ervoor om diagnose specifieke SMRs te gebruiken en te publiceren waarbij het onderliggende casemix correctiemodel gebruik maakt van (1) de Elixhauser comorbiditeit maat, (2) diagnose specifieke, klinisch relevante predictoren, en (3) sterfte kort na ontslag. Verder dienen alleen ziekenhuizen met nagenoeg dezelfde casemix en behandelingsfaciliteiten met elkaar vergeleken te worden. Tenslotte is voorzichtigheid geboden bij de interpretatie van SMRs en dient er actief informatie gegeven te worden over de tekortkomingen van deze getallen.

ACKNOWLEDGEMENTS IN DUTCH
DANKWOORD

'Het maakt niet uit hoe langzaam je gaat, als je maar niet stopt.'
– Confucius

Dit proefschrift is tot stand gekomen dankzij de hulp en inbreng van velen. Een aantal hiervan wil ik in het bijzonder bedanken.

Prof. Dr. C.J. Kalkman. Beste Cor, ontzettend veel dank voor alle inspanningen en tijd die je in mij gestoken hebt. Ik heb bijzonder veel van je geleerd; vooral tijdens jouw 'sabbatical', toen ik bijna wekelijks hele middagen bij jou thuis aan de keukentafel zat en we discussieerden over serieuze, onderzoek-gerelateerde onderwerpen maar ook over minder serieuze zaken in het leven. Ook de vleugel bleef niet onaangeroerd. Jouw invloed als leermeester op mij is zeer groot geweest, niet alleen op onderzoeksgebied maar ook op klinisch vlak.

Prof. Dr. K.G.M. Moons. Beste Carl, hartelijk dank dat jij mijn promotor wilde zijn. Je kennis en enthousiasme voor onderzoek waren zeer stimulerend. Ook je kritische manier van lezen en becommentariëren was uitermate waardevol. Gelukkig werd het aantal uitroeptekens achter het zinnetje '... dit stuk behoeft nog wat aandacht!' over de jaren steeds ietsje minder.

Dr. L.M. Peelen. Beste Linda, ik voelde me door jou in alle opzichten gesteund en gemotiveerd door je heldere inzichten, kennis, efficiëntie en resultaat-gerichtheid. Heel hartelijk dank hiervoor. Overigens denk ik dat we boven-staande quote verschillend interpreteren, gezien het aantal e-mails in mijn inbox met als onderwerp 'wordt het niet weer eens tijd om af te spreken?'.

Dr. H.F. Lingsma. Beste Hester, ik vind het nog steeds ongelooflijk hoe jij telkens binnen een handomdraai begreep wat ik bedoelde. Dank je wel voor de talloze nuttige aanwijzingen.

Ir. A. de Bruin. Beste Agnes, zonder het vertrouwen van jou en daarmee de faciliteiten van het Centraal Bureau voor de Statistiek zou ik dit onderzoek niet gedaan kunnen hebben. Hartelijk dank hiervoor. Ook Janneke Ploemacher en Jan van der Laan wil ik danken voor het beschikbaar stellen van data en meedenken aan mijn onderzoek.

Alle collega-promovendi van de afgelopen 8 jaar. Dank voor jullie input en gezelligheid.

Sectie Anesthesiologie van het Meander MC te Amersfoort. Beste sectie-genoten, mede dankzij jullie interesse ('uiterst boeiend onderwerp collega maar wanneer komt dat feest?'), motivatietechnieken ('Is het nog niet af?'), wijsheden ('Promoveren kost je hooguit je huwelijk') en hulpvaardigheid ('Laat mij eens typen; hoe moeilijk kan dit zijn?') is het boekje nu dan klaar. Hartelijk dank voor jullie flexibele opstelling en bovenal jullie collegialiteit.

Paranimfen Dr. H.T. Faber en Dr. M.B.A. van der Wal. Beste Hubert en Martijn, we zijn sinds onze kennismaking in 2000 een studie geneeskunde, opleiding en PhD traject verder. Toch opvallend dat onze gesprekken nog steeds nul komma nul medisch-inhoudelijke onderwerpen bevatten. Het percentage 'bullshit' onderwerpen hebben we mogelijk wel iets naar beneden gebracht. Dank voor jullie vriendschap.

Ineke Dullaart, de beste schoonmoeder die ik mij kan wensen. Lieve Ineke, dank voor je rotsvaste vertrouwen in mij; alvast het cadeau voor mijn promo-tie geven (inmiddels 2 jaar geleden), is hiervan een van de vele voorbeelden.

Vanessa Dullaart. Lieve Vanessa, dank voor je vriendschap en hulp de afge-lopen jaren. Het feit dat je 'om de hoek' woont is super gezellig maar heeft er wel voor gezorgd dat ik (veel te) veel gebruik van je heb gemaakt om thuis in te springen als ik voor de zoveelste keer 'ietsje pietsje later dan gepland thuis kwam'. Super lief dat je dat keer op keer voor me doet.

Lieve papa en mama, ik dank jullie voor jullie onvoorwaardelijke liefde en steun. Altijd kon ik op jullie rekenen. Eindelijk, jullie hebben er lang op moeten wachten, het is klaar!

Lieve Aurélie en Mathis, de vragen: 'kun je proberen vroeg thuis te komen?' en 'moet je vandaag thuis werken?' staan in mijn geheugen gegrift. I owe you.

Allerliefste Claudia, zonder jou had ik dit boek nooit kunnen voltooien. Zonder jou zou ik überhaupt nooit aan de studie geneeskunde begonnen zijn. Zonder jou zou ik niet kunnen zijn wie ik nu ben. Mét jou heb ik het mooiste wat er is: Aurélie en Mathis. Lieve Clau, dank voor al je liefde, begrip en hulp.

CURRICULUM VITAE

Maurice Emir Pouw was born on September 5, 1976 in Delft, The Netherlands. In 1993 he graduated from high school at the 'Citycollege, St. Franciscus' in Rotterdam. After a study abroad program in Paris, France, he started the study Econometrics in 1994 at the Erasmus University in Rotterdam. In 1999 he received his Master's degree in Financial Econometrics and started working at the ING Bank Netherlands in the department of Policies and Planning (1999-2003). In 2000 he entered medical school at the University of Utrecht.

After his graduation from medical school in 2006 he started his residency in Anesthesiology at the University Medical Center of Utrecht under supervision of Prof. Dr. J.T.A. Knape in June 2007. In 2008 he started his PhD training and the scientific investigation which lead to this thesis. From 2014 on, he combines research with his work as anesthesiologist at the Meander MC in Amersfoort.

Maurice Pouw married Claudia Dullaart in 2006. Together, they have two children, Aurélie (2010), and Mathis (2012). They live in Amsterdam.